PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12Q 1/68, C12N 15/82, A01H 5/00

A1

(11) International Publication Number: **WO 00/18963**

(43) International Publication Date: 6 April 2000 (06.04.00)

(21) International Application Number: PCT/US99/22675

(22) International Filing Date: 30 September 1999 (30.09.99)

(30) Priority Data:

60/102,662	1 October 1998 (01.10.98)	US
60/127,627	1 April 1999 (01.04.99)	US
60/135,608	24 May 1999 (24.05.99)	US

(71) Applicant: MONSANTO COMPANY [US/US]; 800 North Lindbergh Boulevard, St. Louis, MO 63167 (US).

(72) Inventors: DELANNAY, Xavier; St. Louis, MO 63167 (US). CONCIBIDO, Vergel, C.; St. Louis, MO 63167 (US).

(74) Agent: MARSH, David, R.; Howrey & Simon, Box 34, 1299 Pennsylvania Avenue, N.W., Washington, DC 20004-2402 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS FOR BREEDING FOR AND SCREENING OF SOYBEAN PLANTS WITH ENHANCED YIELDS, AND SOYBEAN PLANTS WITH ENHANCED YIELDS

(57) Abstract

The present invention is in the field of plant breeding and genetics, particularly as it pertains to *Glycine max* (soybean). More specifically, the invention relates to quantitative trait loci that are associated with enhanced yield in *Glycine max*, *Glycine max* having such loci and methods for breeding for and screening of *Glycine max* with such loci. The invention further relates to the use of exotic germplasm in a breeding program.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Methods for Breeding for and Screening of Soybean Plants with Enhanced Yields, and Soybean Plants with Enhanced Yields

Cross-Reference to Related Applications

This application claims priority to U.S. provisional application No. 60/102,662 filed October 1, 1998, U.S. provisional application No. 60/127,627 filed April 1, 1999, and U.S. provisional application No. 60/135,608 filed May 24, 1999.

Field of the Invention

The present invention is in the field of plant breeding and genetics, particularly as it pertains to *Glycine max* (soybean). More specifically, the invention relates to alleles of a quantitative trait locus that are associated with enhanced yield in *Glycine max, Glycine max* plants having such alleles and methods for breeding for and screening of *Glycine max* plants with such alleles. The invention further relates to the use of exotic germplasm in a breeding program.

Background of the Invention

The soybean, *Glycine max* (L.) Merril (*Glycine max* or soybean), is one of the major economic crops grown worldwide as a primary source of vegetable oil and protein (Sinclair and Backman, *Compendium of Soybean Diseases*, 3rd Ed. APS Press, St. Paul, MN, p. 106. (1989)). The growing demand for low cholesterol and high fiber diets has also increased soybean's importance as a health food.

Prior to 1940, soybean cultivars were either direct releases of introductions brought from Asia or pure line selections from genetically diverse plant introductions. The soybean plant was primarily used as a hay crop in the early part of the 19th century. Only a few introductions were large-seeded types useful for feed grain and oil production. From the mid 1930's through the 1960's, gains in soybean seed yields were achieved by changing the breeding method from evaluation and selection of introduced germplasm to crossing elite by elite lines. The continuous cycle of cross hybridizing the elite strains selected from the progenies of previous crosses resulted in the modern day cultivars.

Over 10,000 soybean strains have now been introduced into the United States since the early 1900's (Bernard *et al., United States National Germplasm Collections.* In: L.D. Hil (ed.), World Soybean Research, pp. 286-289. Interstate Printers and Publ., Danville, Il. (1976)). A limited number of those introductions form the genetic base of cultivars developed from the hybridization and selection programs (Johnson and Bernard, *The Soybean,* Norman Ed., Academic Press, N.Y. pp. 1-73 (1963)). For

example, in a survey conducted by Specht and Williams, *Genetic Contributions*, Fehr eds. American Soil Association, Wisconsin, pp. 49-73 (1984), for the 136 cultivars released from 1939 to 1989, only 16 different introductions were the source of cytoplasm for 121 of that 136.

Six introductions, 'Mandarin,' 'Manchu,' 'Mandarin' (Ottawa), 'Richland,' 'AK' (Harrow), and 'Mukden,' contributed nearly 70% of the germplasm represented in 136 cultivar releases. To date, modern day cultivars can be traced back from these six soybean strains from southern China. In a study conducted by Cox *et al.*, *Crop Sci. 25:*529-532 (1988), the soybean germplasm is comprised of 90% adapted materials, 9% unadapted, and only 1% from exotic species.

In soybean, the primary gene pool consists of the adapted *Glycine max* (L.) Merrill, (2n = 40) and its wild counterpart *Glycine soja* (2n = 40), which is distributed in China, Japan, peninsular Korea, Taiwan and Russia. *Glycine max* and *Glycine soja* hybridize, produce viable fertile hybrids, and have homologous chromosomes or in some cases differ by a reciprocal translocation or by a paracentric inversion (Hymowitz and Singh, *Taxonomy and Speciation.* In: Soybeans: Improvement, Production, and Uses, Second ed., No. 16. pp. 23-48. J.R. Wilcox (ed.), American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc., Madison, WI. (1988)). Despite the relative ease of crossing the two species together, only a limited number of *Glycine soja* introductions have been screened for economically important traits due to the presence of many traits in *Glycine soja* that are undesirable in an agronomic setting. Only a limited number of publicly released *Glycine max* cultivars in the U.S. and Canada contain a genetic contribution from *Glycine soja* (Bernard *et al., United States National Germplasm Collections.* In: L.D. Hil (ed.), World Soybean Research, pp. 286-289. Interstate Printers and Publ., Danville, Il. (1976)).

It has been reported that there are probably about 2,000 *Glycine soja* plant introductions in the United States (Palmer *et al., Germplasm Diversity within Soybean,* In Soybean: Genetics, Molecular Biology and Biotechnology, Eds. Verma and Shoemaker, CAB International, Wallingford, Oxon, England (1996)). One such plant introduction, *Glycine soja* PI407305, originated from southern China and belongs to maturity group V and is available from United Stated Department of Agriculture Soybean Germplasm Collection, University of Illinois, Urbana - Champaign, USA. PI407305 exhibits a number of undesirable agronomic traits such

as lodging, small seed size, shattering and black seed color (Palmer *et al., Journal of Heredity 76:* 243-247 (1987)). PI407305 also exhibits a degree of genetic divergence that is greater than many other accessions, it produces fertile flowers and crosses with *Glycine max* to produce fertile F_1 plants and existed in a geographical location that is diverse from the ancestors of *Glycine max.*

Marker assisted introgression of traits into plants has been reported. Marker assisted introgression involves the transfer of a chromosome region defined by one or more markers from one germplasm to a second germplasm. An initial step in that process is the localization of the trait by gene mapping. Gene mapping studies to analyze agronomic traits have been reported in many plants including *Glycine max* and *Glycine max* x *Glycine soja.* Gene mapping is the process of determining a gene's position relative to other genes and genetic markers through linkage analysis. The basic principle for linkage mapping is that the closer together two genes are on the chromosome, the more likely they are to be inherited together (Rothwell, *Understanding Genetics. 4th Ed.* Oxford University Press, New York, p. 703 (1988)). Briefly, a cross is made between two genetically compatible but divergent parents relative to traits under study. Genetic markers are then used to follow the segregation of traits under study in the progeny from the cross (often a backcross, F_2, or recombinant inbred population).

Linkage analysis is based on the level at which markers and genes are co-inherited (Rothwell, *Understanding Genetics. 4th Ed.* Oxford University Press, New York, p. 703 (1988)). Statistical tests like chi-square analysis can be used to test the randomness of segregation or linkage (Kochert, *The Rockefeller Foundation International Program on Rice Biotechnology,* University of Georgia Athens, GA, pp. 1-14 (1989)). In linkage mapping, the proportion of recombinant individuals out of the total mapping population provides the information for determining the genetic distance between the loci (Young, *Encyclopedia of Agricultural Science, Vol. 3,* pp. 275-282 (1994)).

Classical mapping studies utilize easily observable, visible traits instead of molecular markers. These visible traits are also known as naked eye polymorphisms. These traits can be morphological like plant height, fruit size, shape and color or physiological like disease response, photoperiod sensitivity or crop maturity. Visible traits are useful and are still in use because they represent actual phenotypes and are easy to score without any specialized lab equipment. By contrast, the other types of

genetic markers are arbitrary loci for use in linkage mapping and often not associated to specific plant phenotypes (Young, *Encyclopedia of Agricultural Science,* Vol. 3, pp. 275-282 (1994)). Many morphological markers cause such large effects on phenotype that they are undesirable in breeding programs. Many other visible traits have the disadvantage of being developmentally regulated (*i.e.*, expressed only certain stages; or at specific tissue and organs). Oftentimes, visible traits mask the effects of linked minor genes making it nearly impossible to identify desirable linkages for selection (Tanksely, *et al., Biotech. 7*:257-264 (1989)).

Although a number of important agronomic characters are controlled by loci having major effects on phenotype, many economically important traits, such as yield and some forms of disease resistance, are quantitative in nature. This type of phenotypic variation in a trait is typically characterized by continuous, normal distribution of phenotypic values in a particular population (Beckmann and Soller, *Oxford Surveys of Plant Molecular Biology, Miffen.* (ed.), Vol. 3, Oxford University Press, UK., pp. 196-250 (1986)). Loci contributing to such genetic variation are thought to be minor genes, as opposed to major genes with large effects that follow a Mendelian pattern of inheritance. Individual loci controlling polygenic traits are also predicted to follow a Mendelian type of inheritance, however the contribution of each locus is expressed as an increase or decrease in the final trait value.

The advent of DNA markers, such as restriction fragment length polymorphism markers (RFLPs), microsatellite markers (SSR), single nucleotide polymorphic markers (SNPs), and random amplified polymorphic DNA markers (RAPDs), allow the resolution of complex, multigenic traits into their individual Mendelian components (Paterson *et al, Nature 335*:721-726 (1988)). A number of applications of RFLPs and other markers have been suggested for plant breeding. Among the potential applications for RFLPs and other markers in plant breeding include: varietal identification (Soller and Beckmann, *Theor. Appl. Genet. 67*:25-33 (1983); Tanksley *et al., Biotech. 7*:257-264 (1989), QTL mapping (Edwards *et al., Genetics 116*:113-115 (1987); Nienhuis *et al., Crop Sci. 27*:797-803 (1987); Osborn *et al., Theor. Appl. Genet. 73*:350-356 (1987); Romero-Severson *et al, Use of RFLPs In Analysis Of Quantitative Trait Loci In Maize, In* Helentjaris and Burr (eds.), pp. 97-102 (1989); Young *et al., Genetics 120*:579-585 (1988); Martin *et al., Science 243*:1725-1728 (1989); Sarfatti *et al., Theor. Appl. Genet. 78*:22-26 (1989); Tanksley *et al., Biotech. 7*:257-264 (1989); Barone *et al., Mol. Gen. Genet. 224*:177-182

(1990); Jung *et al.*, *Theor. Appl. Genet. 79:*663-672 (1990); Keim *et al.*, *Genetics 126:*735-742 (1990); Keim *et al.*, *Theor. Appl. Genet. 79:*465-369 (1990); *Paterson et al.*, *Genetics 124:*735-742 (1990); Martin *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 88:*2336-2340 (1991); Messeguer *et al.*, *Theor. Appl. Genet. 82:*529-536 (1991); Michelmore *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 88:*9828-9832 (1991); Ottaviano *et al.*, *Theor. Appl. Genet. 81:*713-719 (1991); Yu *et al.*, *Theor. Appl. Genet. 81:*471-476 (1991); Diers *et al.*, *Crop Sci. 32:*377-383 (1992); Doebley *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 87:*9888-9892 (1990)), screening genetic resource strains for useful quantitative trait alleles and introgression of these alleles into commercial varieties (Beckmann and Soller, *Theor. Appl. Genet. 67:*35-43 (1983); Tanksley *et al.*, *Biotech. 7:*257-264 (1989)), marker-assisted selection (Tanksley *et al.*, *Biotech. 7:*257-264 (1989)) and map-based cloning (Tanksley *et al.*, *Biotech. 7:*257-264 (1989)). In addition, DNA markers can be used to obtain information about: (1) the number, effect, and chromosomal location of each gene affecting a trait; (2) effects of multiple copies of individual genes (gene dosage); (3) interaction between/among genes controlling a trait (epistasis); (4) whether individual genes affect more than one trait (pleiotropy); and (5) stability of gene function across environments (G x E interactions).

Gene mapping studies associated with QTLs, have focused on agronomic and morphological characters in plants. In maize (*Zea mays* L.), QTLs contributing to heterosis in several quantitative traits have been mapped (Stuber *et al.*, *Genetics 132:*823-839 (1992)), as well as QTLs for heat tolerance (Ottaviano *et al.*, *Theor. Appl. Genet. 81:*713-719 (1991)) and morphological characters distinguishing maize from teosinte (*Zea mays* ssp. *mexicana)* (Doebley *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 87:*9888-9892 (1990)). In tomato, RFLPs have been used in locating and determining effects of QTLs associated with fruit size, pH, soluble solids (Paterson *et al.*, *Genetics 124:*735-742 (1990)) and water use efficiency (Martin *et al*, *Genetics 120:*579-585 (1989)).

Tanksley *et al.* suggested the use of molecular markers to introduce QTLs from exotic germplasm (Tanksley *et al., Theor. Appl. Genet.* 92: 191-203 (1996)). Paterson *et al.*, report the location of putative QTLs in an F_2 population that results from a cross between a domestic tomato strain and an exotic relative (Paterson *et al.*, *Genetics 127:* 181-197 (1991)). The present effort evolved from efforts to locate and introduce traits that enhance agronomical traits into *Glycine max* from *Glycine soja*

introductions. Activities not described by Tanksley *et al., Theor. Appl. Genet.* 92: 191-203 (1996) or Paterson *et al., Genetics 127:* 181-197 (1991).

The present invention provides *Glycine max* plants and methods for producing such plants that address the following difficulties: (A) the level of agronomically detrimental traits associated with *Glycine soja* accessions; (B) the narrow genetic basis of commercial *Glycine max* plants; (C) difficulties associated with the introgression of *Glycine soja* traits into commercial *Glycine max* plants; and (D) difficulties associated with the localization of agronomically desirable traits from *Glycine soja* accessions.

Summary Of The Invention

The present invention includes and provides a *Glycine max* plant having an allele of a quantitative trait locus associated with enhanced yield in the *Glycine max* plant, wherein the allele of the quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

The present invention also provides an elite *Glycine max* plant having an allele of a quantitative trait locus associated with enhanced yield in the elite *Glycine max* plant, wherein the allele of the quantitative trait locus is also located on linkage group U26 of an exotic *Glycine* plant.

The present invention also provides a *Glycine max* plant having a genome, wherein the genome comprises a genetic locus having an allele of a quantitative trait locus genetically linked to the complement of marker nucleic acid molecule U3944117 or its complement.

The present invention also provides a *Glycine max* plant comprising an allele of a quantitative trait locus derived from *Glycine soja* PI407305 or progeny thereof, wherein the quantitative trait locus derived from *Glycine soja* PI407305 or progeny thereof is located on linkage group U26.

The present invention also provides an elite *Glycine max* plant comprising an allele of a quantitative trait locus derived from an exotic *Glycine* plant, wherein the quantitative trait locus is also located on linkage group U26 of *Glycine soja* PI407305.

The present invention also provides a *Glycine max* plant comprising DNA where the DNA has the substantially homologous sequence as DNA found in an allele of a quantitative trait locus derived from *Glycine soja* PI407305 or progeny thereof.

The present invention also provides a *Glycine max* seed from a *Glycine max* plant of the present invention.

The present invention also provides a container of over 40,000 *Glycine max* seeds, wherein over 80% of the seeds have an allele of a quantitative trait locus associated with enhanced yield in the *Glycine max* plant, wherein the allele of a quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

The present invention also provides a *Glycine max* plant, which exhibits an enhanced yield compared to a first parent, the *Glycine max* plant comprising a genome homozygous or heterozygous with respect to a genetic allele that is native to a second parent selected from the group consisting of *Glycine soja* PI407305 and progeny thereof and non-native to a first parent, wherein the first parent is an elite *Glycine max* plant.

The present invention also provides an elite *Glycine max* plant, which exhibits an enhanced yield compared to a first parent, the elite *Glycine max* plant comprising a genome homozygous or heterozygous with respect to a genetic allele that is native to a second parent selected from the group consisting of an exotic *Glycine* plant having an allele of a quantitative trait locus, where the quantitative trait locus is also located on linkage group U26 of *Glycine soja* PI407305.

The present invention also provides a *Glycine max* plant selected for by screening for an enhanced yield in the *Glycine max* plant, the selection comprising interrogating genomic DNA for the presence of a marker molecule that is genetically linked to an allele of a quantitative trait locus associated with enhanced yield in the *Glycine max* plant, wherein the allele of a quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

The present invention also provides an elite *Glycine max* plant selected for by screening for an enhanced yield in the *Glycine max* plant, the selection comprising interrogating genomic DNA for the presence of a marker molecule that is genetically linked to an allele of a quantitative trait locus associated with enhanced yield in an exotic *Glycine* plant, wherein the allele of a quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

The present invention also provides a *Glycine max* plant having a genome, where the genome has a least two polymorphisms capable of being detected by polymorphic markers selected from the group consisting of: Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its

complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement.

The present invention also provides a *Glycine max* seed selected from a *Glycine max* plant by screening for an enhanced yield in the *Glycine max* plant, the selection comprising interrogating genomic DNA for the presence of a marker molecule that is genetically linked to an allele of a quantitative trait locus associated with enhanced yield in the *Glycine max* plant, wherein the allele of the quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

The present invention also provides a substantially purified marker nucleic acid molecule, the marker nucleic acid molecule capable of specifically hybridizing to a second nucleic acid molecule that is U3944117 or its complement.

The present invention also provides a substantially purified marker nucleic acid molecules, the marker nucleic acid molecule capable of specifically hybridizing to a second nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement.

The present invention also provides a substantially purified marker molecule, the marker nucleic acid molecule capable of specifically hybridizing to a region of *Glycine soja* genomic DNA between Satt168 and Satt560.

The present invention also provides a method for the production of an elite *Glycine max* plant having enhanced yield comprising: (A) crossing a *Glycine soja* PI407305 plant or progeny thereof with a *Glycine max* plant to produce a segregating population; (B) screening the segregating population for a member having an allele derived from *Glycine soja* PI407305 plant or progeny thereof that mapped to linkage group U26 of the *Glycine soja* PI407305 plant or progeny thereof, wherein the allele is associated with the enhanced yield in the *Glycine max* plant; and (C) selecting the member for further crossing and selection, wherein the member selected has the allele derived from *Glycine soja* PI407305 plant or progeny thereof that mapped to linkage group U26.

The present invention also provides a substantially purified marker molecule, the marker nucleic acid molecule capable of specifically hybridizing to a region of *Glycine soja* genomic DNA between Satt168 and Satt560.

The present invention also provides a method for the production of an elite *Glycine max* plant having enhanced yield comprising: (A) crossing a *Glycine soja* PI407305 plant or progeny thereof with a *Glycine max* plant to produce a segregating population; (B) screening the segregating population for a member having an allele derived from an exotic *Glycine* plant that also maps to linkage group U26 of the *Glycine soja* PI407305 plant, wherein the allele is associated with the enhanced yield in the *Glycine max* plant; and (C) selecting the member for further crossing and selection, wherein the member selected has the allele derived from said exotic *Glycine* plant.

The present invention also provides a method of introgressing enhanced yield into a *Glycine max* plant comprising using a nucleic acid marker for marker assisted selection of the *Glycine max* plant, the nucleic acid marker complementary to a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of U3944117 or its complement, wherein the source of the enhanced yield is *Glycine soja* PI407305 or progeny thereof, and introgressing the enhanced yield into a *Glycine max* plant.

The present invention also provides a method of introgressing enhanced yield into a *Glycine max* plant comprising using a nucleic acid marker for marker assisted selection of the *Glycine max* plant, the nucleic acid marker complementary to a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of U3944117 or its complement, wherein the source of the enhanced yield is an exotic *Glycine* plant, and introgressing the enhanced yield into a *Glycine max* plant.

The present invention also provides a method of introgressing enhanced yield into a *Glycine max* plant comprising using a nucleic acid marker for marker assisted selection of the *Glycine max* plant, the nucleic acid marker complementary to a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of a nucleic acid marker selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its

complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement, wherein the source of the enhanced yield is *Glycine soja* PI407305 or progeny thereof, and introgressing the enhanced yield into a *Glycine max* plant.

The present invention also provides a method of introgressing enhanced yield into a *Glycine max* plant comprising using a nucleic acid marker for marker assisted selection of the *Glycine max* plant, the nucleic acid marker complementary to a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of a nucleic acid marker selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement, wherein the source of the enhanced yield is an exotic *Glycine* plant, and introgressing the enhanced yield into a *Glycine max* plant.

The present invention also provides a method of introgressing enhanced yield into a *Glycine max* plant comprising using a nucleic acid marker for marker assisted selection of the *Glycine max* plant, the nucleic acid marker complementary to a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant between Satt168 and Satt560, wherein the source of the enhanced yield is *Glycine soja* PI407305 or progeny thereof, and introgressing the enhanced yield into a *Glycine max* plant.

The present invention also provides a method of introgressing enhanced yield into a *Glycine max* plant comprising using a nucleic acid marker for marker assisted selection of said *Glycine max* plant, the nucleic acid marker complementary to a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant between Satt168 and Satt560, wherein the source of the enhanced yield is an exotic *Glycine* plant, and introgressing the enhanced yield into a *Glycine max* plant.

The present invention also provides a method for screening for enhanced yield comprising interrogating genomic DNA for the presence or absence of a marker molecule that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of U3944117 or its complement,

wherein the source of the enhanced yield is *Glycine soja* PI407305 or progeny thereof; and detecting the presence or absence of the marker.

The present invention also provides a method for screening for enhanced yield comprising interrogating genomic DNA for the presence or absence of a marker molecule that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of U3944117 or its complement, wherein the source of the enhanced yield is an exotic *Glycine* plant; and detecting the presence or absence of the marker.

The present invention also provides a method for screening for enhanced yield comprising interrogating genomic DNA for the presence or absence of a marker molecule that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of a nucleic acid marker selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement, wherein the source of the enhanced yield is *Glycine soja* PI407305 or progeny thereof; and detecting the presence or absence of the marker.

The present invention also provides a method for screening for enhanced yield comprising interrogating genomic DNA for the presence or absence of a marker molecule that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of a nucleic acid marker selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement, wherein the source of the enhanced yield is an exotic *Glycine* plant; and detecting the presence or absence of the marker.

The present invention provides a method for screening for enhanced yield comprising interrogating genomic DNA for the presence or absence of a marker molecule that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant between Satt168 and Satt560, wherein the source of

the enhanced yield is *Glycine soja* PI407305 or progeny thereof; and detecting the presence or absence of the marker.

The present invention provides a method for screening for enhanced yield comprising interrogating genomic DNA for the presence or absence of a marker molecule that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant between Satt168 and Satt560, wherein the source of the enhanced yield is an exotic *Glycine* plant; and detecting the presence or absence of the marker.

The present invention also provides a method for determining the likelihood of a quantitative trait allele for enhanced yield in a *Glycine max* plant comprising the steps of: (A) obtaining genomic DNA from the plant; (B) detecting a marker molecule; wherein the marker molecule specifically hybridizes with a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of U3944117 or its complement and (C) determining the presence or absence of the marker molecule, wherein the presence or absence of the marker molecule is indicative of the quantitative trait allele for enhanced yield.

The present invention also provides a method for determining the likelihood of a quantitative trait allele for enhanced yield in a *Glycine max* plant comprising the steps of: (A) obtaining genomic DNA from the plant; (B) detecting a marker molecule; wherein the marker molecule specifically hybridizes with a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of a nucleic acid marker selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement and (C) determining the presence or absence of the marker molecule, wherein the presence or absence of the marker molecule is indicative of the quantitative trait allele for enhanced yield.

The present invention provides a method for determining the likelihood of a quantitative trait allele for enhanced yield in a *Glycine max* plant comprising the steps of: (A) obtaining genomic DNA from the plant; (B) detecting a marker molecule, wherein said marker molecule specifically hybridizes with a nucleic acid sequence

that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant between and Satt168 and Satt560; and (C) determining the presence or absence of the marker molecule, wherein the presence or absence of the marker molecule is indicative of the quantitative trait allele for enhanced yield.

The present invention also provides a method for determining the probability that a plant has a quantitative trait allele for enhanced yield: (A) detecting the presence or absence of a polymorphism genetically or physically linked to a quantitative trait allele for enhanced yield, wherein the polymorphism is located on linkage group U26 of a *Glycine max* plant within 50 cM of U3944117 or its complement; and (B) determining the probability that the plant has the quantitative trait allele for enhanced yield.

The present invention also provides a method for determining the probability that a plant has quantitative trait alleles for enhanced yield: (A) detecting the presence or absence of a polymorphism genetically or physically linked to quantitative trait alleles for enhanced yield, wherein the polymorphism is located on linkage group U26 of a *Glycine max* plant within 50 cM of a nucleic acid marker selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement; and (B) determining the probability that the plant has the quantitative trait alleles for enhanced yield.

The present invention also provides a method for determining a genomic polymorphism in a plant that is predictive of an enhanced yield comprising the steps: (A) incubating a marker nucleic acid molecule, under conditions permitting nucleic acid hybridization, and a complementary nucleic acid molecule obtained from the plant, the marker nucleic acid molecule selected from the group consisting of a marker nucleic acid molecule that specifically hybridizes to Satt168 or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt416 or its complement, a marker nucleic acid molecule that specifically hybridizes to Sat_083 or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt474 or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt122 or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt556 or its complement, a marker nucleic acid molecule that specifically

hybridizes to Sct_094 or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt272 or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt020 or its complement, a marker nucleic acid molecule that specifically hybridizes to U3944117 or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt066 or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt534 or its complement, and a marker nucleic acid molecule that specifically hybridizes to Satt560 or its complement; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant; and (C) detecting the presence of the polymorphism.

The present invention also provides a method of determining an association between a polymorphism and a plant trait comprising: (A) hybridizing a nucleic acid molecule specific for the polymorphism to genetic material of a plant, wherein the nucleic acid molecule or complement thereof is selected from the group consisting of a nucleic acid molecule that is complementary to a nucleic acid sequence that is genetically linked to a quantitative trait locus in a region between and including a nucleic acid sequence that specifically hybridizes to a region between and including nucleic acid marker U3944117 and within 50 cM of U3944117 or its complement on linkage group U26; and (B) calculating the degree of association between the polymorphism and the plant trait.

The present invention provides a method of determining an association between a polymorphism and a plant trait comprising: (A) hybridizing a nucleic acid molecule specific for the polymorphism to genetic material of a plant, wherein the nucleic acid molecule or complement thereof is selected from the group consisting of a nucleic acid molecule that is complementary to a nucleic acid sequence that is genetically linked to a quantitative trait locus in a region between and including a nucleic acid sequence that specifically hybridizes to a region between Satt168 and Satt560; and (B)calculating the degree of association between the polymorphism and the plant trait.

Description Of The Figures

Figure 1 diagrammatically sets forth the location of marker molecules on chromosome U26 (B2). On this map Satt168 is about 2.9 cM from Satt416, which is about 4.5 cM from Sat_083, which is about 0.0 cM from Satt474, which is about 0.0 cM from Satt122, which is about 0.0 cM from Satt556, which is about 0.5 cM from

Sct_094, which is about 1.0 cM from Satt272, which is about 0.0 cM from Satt020, which is about 0.8 cM from U3944117, which is about 2.4 cM from Satt066, which is about 4.5 cM from Satt534, which is about 8.1 cM from Satt560. The LOD score for the peak where the yield QTL is located is about 3.99.

Description of the Sequence Listings

The following sequence listings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these sequences in combination with the detailed description presented herein.

SEQ ID NO. 1. U3944117
SEQ ID NO. 2. U3944117b
SEQ ID NO. 3. SATT168
SEQ ID NO. 4. SATT416
SEQ ID NO. 5. SAT_083
SEQ ID NO. 6. SATT474
SEQ ID NO. 7. SATT122
SEQ ID NO. 8. SATT556
SEQ ID NO. 9. SCT_094
SEQ ID NO. 10. SATT272
SEQ ID NO. 11 SATT020
SEQ ID NO. 12. SATT066
SEQ ID NO. 13. SATT534
SEQ ID NO. 14. SATT560
SEQ ID NO. 15. E39 primer
SEQ ID NO. 16. M44 primer
SEQ ID NO. 17. 168 forward primer
SEQ ID NO. 18. 168 reverse primer
SEQ ID NO. 19. 416 forward primer
SEQ ID NO. 20. 416 reverse primer
SEQ ID NO. 21. 083 forward primer
SEQ ID NO. 22. 083 reverse primer
SEQ ID NO. 23. 474 forward primer
SEQ ID NO. 24. 474 reverse primer

SEQ ID NO. 25. 122 forward primer
SEQ ID NO. 26. 122 reverse primer
SEQ ID NO. 27. 556 forward primer
SEQ ID NO. 28. 556 reverse primer
SEQ ID NO. 29. 094 forward primer
SEQ ID NO. 30. 094 reverse primer
SEQ ID NO. 31. 272 forward primer
SEQ ID NO. 32. 272 reverse primer
SEQ ID NO. 33. 020 forward primer
SEQ ID NO. 34. 020 reverse primer
SEQ ID NO. 35. 066 forward primer
SEQ ID NO. 36. 066 reverse primer
SEQ ID NO. 37. 534 forward primer
SEQ ID NO. 38. 534 reverse primer
SEQ ID NO. 39. 560 forward primer
SEQ ID NO. 40. 560 reverse primer

Detailed Description Of The Invention

The present invention provides a *Glycine max* plant having an allele of a quantitative trait locus associated with enhanced yield in the *Glycine max* plant, where the allele of a quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

A *Glycine max* plant of the present invention is any *Glycine max* plant. In a preferred embodiment, a *Glycine max* plant of the present invention is an elite plant. An "elite line" is any line that has resulted from breeding and selection for superior agronomic performance. Examples of elite lines are lines that are commercially available to farmers or soybean breeders such as HARTZ™ variety H4994, HARTZ™ variety H5218, HARTZ™ variety H5350, HARTZ™ variety H5545, HARTZ™ variety H5050, HARTZ™ variety H5454, HARTZ™ variety H5233, HARTZ™ variety H5488, HARTZ™ variety HLA572, HARTZ™ variety H6200, HARTZ™ variety H6104, HARTZ™ variety H6255, HARTZ™ variety H6586, HARTZ™ variety H6191, HARTZ™ variety H7440, HARTZ™ variety H4452 Roundup Ready™, HARTZ™ variety H4994 Roundup Ready™, HARTZ™ variety H4988 Roundup Ready™, HARTZ™ variety H5000 Roundup Ready™, HARTZ™ variety H5147 Roundup Ready™, HARTZ™ variety H5247 Roundup Ready™, HARTZ™

variety H5350 Roundup Ready™, HARTZ™ variety H5545 Roundup Ready™, HARTZ™ variety H5855 Roundup Ready™, HARTZ™ variety H5088 Roundup Ready™, HARTZ™ variety H5164 Roundup Ready™, HARTZ™ variety H5361 Roundup Ready™, HARTZ™ variety H5566 Roundup Ready™, HARTZ™ variety H5181 Roundup Ready™, HARTZ™ variety H5889 Roundup Ready™, HARTZ™ variety H5999 Roundup Ready™, HARTZ™ variety H6013 Roundup Ready™, HARTZ™ variety H6255 Roundup Ready™, HARTZ™ variety H6454 Roundup Ready™, HARTZ™ variety H6686 Roundup Ready™, HARTZ™ variety H7152 Roundup Ready™, HARTZ™ variety H7550 Roundup Ready™, HARTZ™ variety H8001 Roundup Ready™ (HARTZ SEED, Stuttgart, Arkansas, U.S.A.); A0868, AG0901, A1553, A1900, AG1901, A1923, A2069, AG2101, AG2201, A2247, AG2301, A2304, A2396, AG2401, AG2501, A2506, A2553, AG2701, AG2702, A2704, A2833, A2869, AG2901, AG2902, AG3001, AG3002, A3204, A3237, A3244, AG3301, AG3302, A3404, A3469, AG3502, A3559, AG3601, AG3701, AG3704, AG3750, A3834, AG3901, A3904, A4045 AG4301, A4341, AG4401, AG4501, AG4601, AG4602, A4604, AG4702, AG4901, A4922, AG5401, A5547, AG5602, A5704, AG5801, AG5901, A5944, A5959, AG6101, QR4459 and QP4544 (Asgrow Seeds, Des Moines, Iowa, U.S.A.); DeKalb variety CX445 (DeKalb, Illinois). An elite plant is any plant from an elite line.

The quantitative trait locus of the present invention may be introduced into an elite *Glycine max* transgene that contains one or more genes for herbicide resistance, increased yield, insect control, fungal disease resistance, virus resistance, nematode resistance, bacterial disease resistance, mycoplasma disease resistance, modified oils production, high protein production, germination and seedling growth control, enhanced animal and human nutrition, low raffinose, environmental stress resistant, increased digestibility, industrial enzymes, pharmaceuticals, improved processing traits, nitrogen fixation, hybrid seed production, among others.

In a further preferred embodiment, the nuclear genetic contribution of *Glycine soja* to a *Glycine max* of the present invention is less than about 25%. In a more preferred embodiment, the nuclear genetic contribution of *Glycine soja* to a *Glycine max* of the present invention is less than about 12.5%. In an even more preferred embodiment, the nuclear genetic contribution of *Glycine soja* to a *Glycine max* of the present invention is less than about 6.25%. The *Glycine soja* genetic contribution in a *Glycine max* plant of the present invention can be reduced by backcrossing the

progeny of a *Glycine max* x *Glycine soja* cross (or progeny thereof) with, for example, a *Glycine max* recurrent parent.

In a further preferred embodiment, the nuclear genetic contribution of an exotic *Glycine* to a *Glycine max* of the present invention is less than about 25%. In a more preferred embodiment, the nuclear genetic contribution of an exotic *Glycine* to a *Glycine max* of the present invention is less than about 12.5%. In an even more preferred embodiment, the nuclear genetic contribution of an exotic *Glycine* to a *Glycine max* of the present invention is less than about 6.25%. The an exotic *Glycine* genetic contribution in a *Glycine max* plant of the present invention can be reduced by backcrossing the progeny of a *Glycine max* x an exotic *Glycine* cross (or progeny thereof) with, for example, a *Glycine max* recurrent parent. It is further understood that a *Glycine max* plant of the present invention may exhibit the characteristics of any maturity group.

A number of molecular genetic maps of *Glycine* have been reported (Mansur *et al., Crop Sci. 36:* 1327-1336 (1996); Shoemaker *et al., Genetics 144:* 329-338 (1996); Shoemaker *et al., Crop Science 32:* 1091-1098 (1992); Shoemaker *et al., Crop Science 35:* 436-446 (1995); Tinley and Rafalski, *J. Cell Biochem. Suppl. 14E*: 291 (1990)). *Glycine max, Glycine soja* and *Glycine max* x. *Glycine soja* share linkage groups (Shoemaker *et al., Genetics 144:* 329-338 (1996)). As used herein, reference to the U26 linkage group of *Glycine soja* refers to the linkage group that corresponds to U26 linkage group from the genetic map of *Glycine max* (Mansur *et al., Crop Sci. 36:* 1327-1336 (1996) and B2 linkage group *Glycine max* x. *Glycine soja* (Shoemaker *et al., Genetics 144:* 329-336 (1996)) that is present in *Glycine soja* (Soybase, an Agricultural Research Service, United States Department of Agriculture (http://129.186.26.940/ and USDA - Agricultural Research Service: http://www.ars.usda.gov/)).

An allele of a quantitative trait locus can, of course, comprise multiple genes or other genetic factors even within a contiguous genomic region or linkage group. As used herein, an allele of a quantitative trait locus can therefore encompasses more than one gene or other genetic factor where each individual gene or genetic component is also capable of exhibiting allelic variation and where each gene or genetic factor also has a phenotypic effect on the quantitative trait in question. In an embodiment of the present invention the allele of a quantitative trait locus comprises one or more genes or other genetic factors that are also capable of exhibiting allelic

variation. The use of the term "an allele of a quantitative trait locus" is thus not intended to exclude a quantitative trait locus that comprises more than one gene or other genetic factor.

It is further understood that a *Glycine soja* plant may be any *Glycine soja* plant having an allele of a quantitative trait locus that is associated with enhanced yield when the allele is present in a *Glycine max* plant. In a preferred embodiment, an allele of a quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant. In an even more preferred embodiment, an allele of the quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant is genetically linked to a complement of a marker nucleic acid, where the marker nucleic acid molecule is selected from the group consisting of a marker nucleic acid molecule in a region between and including marker U3944117 and within 50 cM of U3944117 or its complement on linkage group U26. A preferred *Glycine soja* plant introduction for use in conjunction with any aspect of the present invention is PI407305.

The present invention includes and provides an elite *Glycine max* plant having an allele of a quantitative trait locus associated with enhanced yield in the elite *Glycine max* plant, wherein the allele of the quantitative trait locus is also located on linkage group U26 of an exotic *Glycine* plant. As used herein, an exotic is an non-elite *Glycine* species. In a preferred embodiment, the non-elite *Glycine* species is a species where less than 50%, more preferably 75%, of the germplasm genetic composition is derived from the following six introductions: Mandarin, Manchu, Mandarin (Ottawa), Richland, AK (Harrow) and Mukden.

In another preferred embodiment, an allele of a quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant is genetically linked to a complement of a marker nucleic acid, where the marker nucleic acid molecule is selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement.

In a more preferred embodiment, an allele of a quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant is genetically linked to 2 or more, even more preferably 3 or more, 4 or more 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more of the markers selected from

the the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement.

In a more preferred embodiment, an allele of a quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant is genetically linked to 2 or more, even more preferably 3 or more, 4 or more 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more of the markers selected from the the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement and where the markers have approximately the following relative locations: Satt168 is about 2.9 cM from Satt416, which is about 4.5 cM from Sat_083, which is about 0.0 cM from Satt474, which is about 0.0 cM from Satt122, which is about 0.0 cM from Satt556, which is about 0.5 cM from Sct_094, which is about 1.0 cM from Satt272, which is about 0.0 cM from Satt020, which is about 0.8 cM from U3944117, which is about 2.4 cM from Satt066, which is about 4.5 cM from Satt534, which is about 8.1 cM from Satt560.

In another preferred embodiment, an allele of a quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant in a region between and including Satt168 and Satt560. In another preferred embodiment, an allele of the quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant in a region between and including Satt416 and Satt534. In another preferred embodiment, an allele of the quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant in a region between and including Sat_083 and Satt066. In another preferred embodiment, an allele of the quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant in a region between and including Satt474 and Satt066. In another preferred embodiment, an allele of the quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant in a region between and including Satt122 and Satt066. In another preferred embodiment, an allele of the quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant in a region between and including Satt556 and

Satt066. In another preferred embodiment, an allele of the quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant in a region between and including Sct_094 and Satt066. In another preferred embodiment, an allele of the quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant in a region between and including Satt272 and Satt066. In another preferred embodiment, an allele of the quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant in a region between and including Satt020 and Satt066. In another preferred embodiment, an allele of the quantitative trait locus that is also located on linkage group U26 of the *Glycine soja* plant in a region between and including U3944117 and Satt066.

In another preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of *Glycine soja* plant and the allele is also located between about 0 and about 50 centimorgans (cM) from U3944117 or its complement. In another even more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of a *Glycine soja* plant and the allele is also located between about 0 and about 40 centimorgans from U3944117 or its complement. In another even more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of a *Glycine soja* plant and the allele is also located between about 0 and about 25 centimorgans from U3944117 or its complement. In another even more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of a *Glycine soja* plant and the allele is also located between about 0 and about 10 centimorgans from U3944117 or its complement. In another even more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of a *Glycine soja* plant and the allele is also located between about 0 and about 5 centimorgans from U3944117 or its complement. In another even more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of a *Glycine soja* plant and the allele is also located between about 0 and about 3 centimorgans from U3944117 or its complement.

In another preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of

Glycine soja plant and the allele is also located between about 0 and about 50 centimorgans (cM) from a marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement. In another even more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of a *Glycine soja* plant and the allele is also located between about 0 and about 40 centimorgans from a marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement. In another even more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of a *Glycine soja* plant and the allele is also located between about 0 and about 25 centimorgans from a marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement. In another even more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of a *Glycine soja* plant and the allele is also located between about 0 and about 10 centimorgans from a marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement. In another even more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on

linkage group U26 of a *Glycine soja* plant and the allele is also located between about 0 and about 5 centimorgans from a marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement. In another even more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is also located on linkage group U26 of a *Glycine soja* plant and the allele is also located between about 0 and about 3 centimorgans from from a marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement.

In another embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is genetically linked to the marker nucleic acid molecule U3944117 or its complement, where the marker nucleic acid molecule exhibits a LOD score of greater than 2.0, as judged by interval mapping, for enhanced yield, preferably where the marker nucleic acid molecule exhibits a LOD score of greater than 3.0, as judged by interval mapping, for enhanced yield, more preferably where the marker nucleic acid molecule exhibits a LOD score of greater than 3.5, as judged by interval mapping, for enhanced yield and even more preferably where the marker nucleic acid molecule exhibits a LOD score of about 4.0, as judged by interval mapping, for enhanced yield.

In another embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is genetically linked to the marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement, where the marker nucleic acid molecule exhibits a LOD score of greater than 2.0, as judged by

interval mapping, for enhanced yield, preferably where the marker nucleic acid molecule exhibits a LOD score of greater than 3.0, as judged by interval mapping, for enhanced yield, more preferably where the marker nucleic acid molecule exhibits a LOD score of greater than 3.5, as judged by interval mapping, for enhanced yield and even more preferably where the marker nucleic acid molecule exhibits a LOD score of about 4.0, as judged by interval mapping, for enhanced yield.

As used herein, allele is one of several alternative forms of a gene occupying a given locus on a chromosome. When all the alleles present at a given locus on a chromosome are the same that plant is homozygous at that locus. If the alleles present at a given locus on a chromosome differ that plant is heterozygous at that locus.

In an embodiment, a *Glycine max* plant of the present invention exhibits an enhanced yield as measured by dry seed weight. The enhanced yield is measured as dry seed weight at 13% moisture content in comparison to a *Glycine max* plant of a similar genetic background grown under similar conditions but whose genetic makeup lacks the alleles of a quantitative trait locus associated with enhanced yield in the *Glycine max* plant, where the alleles of a quantitative trait locus are also located on linkage group U26 of a *Glycine soja* plant. In an embodiment the enhanced yield results in a greater than 2% increase in average dry seed weight. In a preferred embodiment the enhanced yield results in a greater than 4% increase in average dry seed weight. In a more preferred embodiment the enhanced yield results in a greater than 5% increase in average dry seed weight. In an even more preferred embodiment the enhanced yield results in a greater than 10% increase in average dry seed weight. In an even more preferred embodiment the enhanced yield results in a greater than 12% increase in average dry seed weight. In a particularly preferred embodiment the enhanced yield results in a greater than 14% or greater than 18% increase in average dry seed weight.

Many agronomic traits can affect yield. These include, without limitation, plant height, pod number, pod position on the plant, number of internodes, incidence of pod shatter, grain size, efficiency of nodulation and nitrogen fixation, efficiency of nutrient assimilation, resistance to biotic and abiotic stress, carbon assimilation, plant architecture, resistance to lodging, percent seed germination, seedling vigor, and juvenile traits. In an embodiment, a *Glycine max* plant of the present invention exhibits an enhanced trait that is a component of yield.

In another embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is genetically linked to the marker nucleic acid molecule U3944117 or its complement, where the association between the marker nucleic acid molecule and an enhanced yield exhibits a P-value of less than 0.01 for the probability of that association being by chance. In a preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is genetically linked to the marker nucleic acid molecule U3944117 or its complement, where the association between the marker nucleic acid molecule and an enhanced yield exhibits a P-value of less than 0.001 for the probability of that association being by chance. In a more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is genetically linked to the marker nucleic acid molecule U3944117 or its complement, where the association between the marker nucleic acid molecule and an enhanced yield exhibits a P-value of less than 0.0001 for the probability of that association being by chance.

In another embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is genetically linked to a marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement, where the association between the marker nucleic acid molecule and an enhanced yield exhibits a P-value of less than 0.01 for the probability of that association being by chance. In a preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is genetically linked to a marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement, where the association between the marker nucleic acid molecule and an enhanced yield exhibits a P-value of less than 0.001 for the probability of that association being by chance. In a more preferred embodiment, a *Glycine max* plant of the present invention has an allele of a quantitative trait locus that is genetically linked to a marker nucleic acid

molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement, where the association between the marker nucleic acid molecule and an enhanced yield exhibits a P-value of less than 0.0001 for the probability of that association being by chance.

In addition, an allele of a quantitative trait locus associated with enhanced yield in *Glycine max* can be associated with any linkage group in *Glycine max.* In a preferred embodiment an allele of a quantitative trait locus associated with enhanced yield in *Glycine max* is located on linkage group U26 of *Glycine max.*

The present invention also provides for a *Glycine max* plant having a genome, where the genome has a genetic locus of a quantitative trait locus having an allele genetically linked to the marker nucleic acid molecule U3944117 or its complement, where the allele is also found in *Glycine soja.*

The present invention also provides for a *Glycine max* plant having a genome, where the genome has a genetic locus of a quantitative trait locus having an allele genetically linked to the marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement, where the allele is also found in *Glycine soja.*

The present invention also provides a *Glycine max* plant having an allele of a quantitative trait locus derived from *Glycine soja* PI407305 or progeny thereof, where the quantitative trait locus derived from *Glycine soja* PI407305 or progeny thereof is located on linkage group U26.

The present invention also provides a *Glycine max* plant having an allele of a quantitative trait locus derived from an exotic *Glycine* plant, where the quantitative trait locus derived from an exotic *Glycine* plant is located on linkage group U26.

Heterogeneity can exist in any *Glycine soja* accession and specifically that heterogeneity may exist in *Glycine soja* PI407305. It is further understood that in light of the current disclosure, *Glycine soja* PI407305 having an allele of a

quantitative trait locus associated with enhanced yield in a *Glycine max* plant can be screened for using one or more the techniques described herein or known in the art. In a preferred embodiment single seed selection from the segregating progeny of PI407305 is used in a backcross with a commercial *Glycine max* lines such as HS-1 and A3244. The presence or absence of alleles from *Glycine soja* PI407305 can, for example, be determined in the BC_2F_4 generation.

As used herein, the progeny of *Glycine soja* or *Glycine soja* PI407305 include not only, without limitation, the products of any cross (be it a backcross or otherwise) between that *Glycine soja* plant, but all progeny whose pedigree traces back to the original cross. Specifically, without limitation, such progeny includes plants that have 12.5% or less genetic material derived from *Glycine soja*. As used herein, a second plant is derived from a first plant if the second plant's pedigree includes the first plant.

The present invention also provides a *Glycine max* plant, which exhibits an enhanced yield compared to a first parent, the *Glycine max* plant having a genome homozygous or heterozygous with respect to a genetic allele that is native to a second parent selected from the group consisting of *Glycine soja* PI407305 and progeny thereof and non-native to a first parent, where the first parent is an elite *Glycine max* plant.

Moreover, the present invention also provides a *Glycine max* plant comprising an allele of a quantitative trait locus derived from an exotic *Glycine* plant, wherein the quantitative trait locus is also located on linkage group U26 of *Glycine soja* PI407305.

Furthermore, the present invention provides a method for the production of an elite *Glycine max* plant having enhanced yield comprising: (A) crossing a *Glycine soja* PI407305 plant or progeny thereof with a *Glycine max* plant to produce a segregating population; (B) screening the segregating population for a member having an allele derived from *Glycine soja* PI407305 plant or progeny thereof that mapped to linkage group U26 of the *Glycine soja* PI407305 plant or progeny thereof, where the allele is associated with the enhanced yield in the *Glycine max* plant; and (C) selecting the member for further crossing and selection, wherein the member selected has the allele derived from *Glycine soja* PI407305 plant or progeny thereof that mapped to linkage group U26.

The present invention further provides a method of introgressing enhanced yield into a *Glycine max* plant comprising using a nucleic acid marker for marker-assisted selection of the *Glycine max* plant, the nucleic acid marker complementary to

a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant, where the source of the enhanced yield is *Glycine soja* PI407305 or progeny thereof.

The present invention also provides a method of introgressing enhanced yield into a *Glycine max* plant comprising using a nucleic acid marker for marker assisted selection of the *Glycine max* plant, the nucleic acid marker complementary to a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant within 50 cM of U3944117 or its complement, wherein the source of the enhanced yield is an exotic *Glycine* plant, and introgressing the enhanced yield into a *Glycine max* plant.

Plants of the present invention can be part of or generated from a breeding program. The choice of breeding method depends on the mode of plant reproduction, the heritability of the trait(s) being improved, and the type of cultivar used commercially (*e.g.*, F_1 hybrid cultivar, pureline cultivar, etc). Selected, non-limiting approaches, for breeding the plants of the present invention are set forth below. A breeding program can be enhanced using marker assisted selection of the progeny of any cross. It is further understood that any commercial and non-commercial cultivars can be utilized in a breeding program. Factors such as, for example, emergence vigor, vegetative vigor, stress tolerance, disease resistance, branching, flowering, seed set, seed size, seed density, standability, and threshability etc. will generally dictate the choice.

For highly heritable traits, a choice of superior individual plants evaluated at a single location will be effective, whereas for traits with low heritability, selection should be based on mean values obtained from replicated evaluations of families of related plants. Popular selection methods commonly include pedigree selection, modified pedigree selection, mass selection, and recurrent selection. In a preferred embodiment a backcross or recurrent breeding program is undertaken.

The complexity of inheritance influences choice of the breeding method. Backcross breeding can be used to transfer one or a few favorable genes for a highly heritable trait into a desirable cultivar. This approach has been used extensively for breeding disease-resistant cultivars. Various recurrent selection techniques are used to improve quantitatively inherited traits controlled by numerous genes. The use of recurrent selection in self-pollinating crops depends on the ease of pollination, the

frequency of successful hybrids from each pollination, and the number of hybrid offspring from each successful cross.

Breeding lines can be tested and compared to appropriate standards in environments representative of the commercial target area(s) for two or more generations. The best lines are candidates for new commercial cultivars; those still deficient in traits may be used as parents to produce new populations for further selection.

One method of identifying a superior plant is to observe its performance relative to other experimental plants and to a widely grown standard cultivar. If a single observation is inconclusive, replicated observations can provide a better estimate of its genetic worth. A breeder can select and cross two or more parental lines, followed by repeated selfing and selection, producing many new genetic combinations.

The development of new soybean cultivars requires the development and selection of soybean varieties, the crossing of these varieties and selection of superior hybrid crosses. The hybrid seed can be produced by manual crosses between selected male-fertile parents or by using male sterility systems. Hybrids are selected for certain single gene traits such as pod color, flower color, seed yield, pubescence color or herbicide resistance which indicate that the seed is truly a hybrid. Additional data on parental lines, as well as the phenotype of the hybrid, influence the breeder's decision whether to continue with the specific hybrid cross.

Pedigree breeding and recurrent selection breeding methods can be used to develop cultivars from breeding populations. Breeding programs combine desirable traits from two or more cultivars or various broad-based sources into breeding pools from which cultivars are developed by selfing and selection of desired phenotypes. New cultivars can be evaluated to determine which have commercial potential.

Pedigree breeding is used commonly for the improvement of self-pollinating crops. Two parents who possess favorable, complementary traits are crossed to produce an F_1. An F_2 population is produced by selfing one or several F_1's. Selection of the best individuals in the best families is selected. Replicated testing of families can begin in the F_4 generation to improve the effectiveness of selection for traits with low heritability. At an advanced stage of inbreeding (*i.e.*, F_6 and F_7), the best lines or mixtures of phenotypically similar lines are tested for potential release as new cultivars.

Backcross breeding has been used to transfer genes for a simply inherited, highly heritable trait into a desirable homozygous cultivar or inbred line, which is the recurrent parent. The source of the trait to be transferred is called the donor parent. The resulting plant is expected to have the attributes of the recurrent parent (*e.g.*, cultivar) and the desirable trait transferred from the donor parent. After the initial cross, individuals possessing the phenotype of the donor parent are selected and repeatedly crossed (backcrossed) to the recurrent parent. The resulting parent is expected to have the attributes of the recurrent parent (*e.g.*, cultivar) and the desirable trait transferred from the donor parent.

The single-seed descent procedure in the strict sense refers to planting a segregating population, harvesting a sample of one seed per plant, and using the one-seed sample to plant the next generation. When the population has been advanced from the F_2 to the desired level of inbreeding, the plants from which lines are derived will each trace to different F_2 individuals. The number of plants in a population declines each generation due to failure of some seeds to germinate or some plants to produce at least one seed. As a result, not all of the F_2 plants originally sampled in the population will be represented by a progeny when generation advance is completed.

In a multiple-seed procedure, soybean breeders commonly harvest one or more pods from each plant in a population and thresh them together to form a bulk. Part of the bulk is used to plant the next generation and part is put in reserve. The procedure has been referred to as modified single-seed descent or the pod-bulk technique.

The multiple-seed procedure has been used to save labor at harvest. It is considerably faster to thresh pods with a machine than to remove one seed from each by hand for the single-seed procedure. The multiple-seed procedure also makes it possible to plant the same number of seed of a population each generation of inbreeding.

Descriptions of other breeding methods that are commonly used for different traits and crops can be found in one of several reference books (*e.g.*, Fehr, *Principles of Cultivar Development* Vol. 1, pp. 2-3 (1987)).

The present invention also provides for parts of the plants of the present invention. Plant parts, without limitation, include seed, endosperm, ovule and pollen. In a particularly preferred embodiment of the present invention, the plant part is a seed.

Moreover, the present invention also provides for a container having more than 40,000 *Glycine max* seeds where over 40% of the seeds are from plants of the present invention. The present invention also provides for a container having more than 80,000 *Glycine max* seeds where over 40% of the seeds are from plants of the present invention.

In a preferred embodiment, the present invention also provides for a container having more than 40,000 *Glycine max* seeds where over 60% of the seeds are from plants of the present invention. In another preferred embodiment, the present invention also provides for a container having more than 80,000 *Glycine max* seeds where over 60% of the seeds are from plants of the present invention. In an even more preferred embodiment, the present invention also provides for a container having more than 40,000 *Glycine max* seeds where over 80% of the seeds are from plants of the present invention. In another even more preferred embodiment, the present invention also provides for a container having more than 80,000 *Glycine max* seeds where over 80% of the seeds are from plants of the present invention. In a further even more preferred embodiment, the present invention also provides for a container having more than 40,000 *Glycine max* seeds where over 90% of the seeds are from plants of the present invention. In another preferred embodiment, the present invention also provides for a container having more than 80,000 *Glycine max* seeds where over 90% of the seeds are from plants of the present invention.

Moreover, the present invention also provides for a container having more than 25 lbs. of *Glycine max* seeds where over 40% of the seeds are from plants of the present invention. The present invention also provides for a container having more than 40lbs. of *Glycine max* seeds where over 40% of the seeds are from plants of the present invention. In a preferred embodiment, the present invention also provides for a container having more than 25lbs. of *Glycine max* seeds where over 60% of the seeds are from plants of the present invention. In another preferred embodiment, the present invention also provides for a container having more than 40lbs. of *Glycine max* seeds where over 60% of the seeds are from plants of the present invention. In an even more preferred embodiment, the present invention also provides for a container having more than 25lbs. of *Glycine max* seeds where over 80% of the seeds are from plants of the present invention. In another even more preferred embodiment, the present invention also provides for a container having more than 40lbs.of *Glycine max* seeds where over 80% of the seeds are from plants of the present invention. In a

further even more preferred embodiment, the present invention also provides for a container having more than 25lbs. of *Glycine max* seeds where over 90% of the seeds are from plants of the present invention. In another preferred embodiment, the present invention also provides for a container having more than 40lbs. of *Glycine max* seeds where over 90% of the seeds are from plants of the present invention.

Plants or parts thereof of the present invention may be grown in culture and regenerated. Methods for the regeneration of *Glycine max* plants from various tissue types and methods for the tissue culture of *Glycine max* are known in the art (See, for example, Widholm *et al., In Vitro Selection and Culture-induced Variation in Soybean,* In Soybean: Genetics, Molecular Biology and Biotechnology, Eds. Verma and Shoemaker, CAB International, Wallingford, Oxon, England (1996)). Regeneration techniques for plants such as *Glycine max* can use as the starting material a variety of tissue or cell types. With *Glycine max* in particular, regeneration processes have been developed that begin with certain differentiated tissue types such as meristems, Cartha *et al., Can. J. Bot. 59*:1671-1679 (1981), hypocotyl sections, Cameya *et al., Plant Science Letters 21:* 289-294 (1981), and stem node segments, Saka *et al., Plant Science Letters, 19:* 193-201 (1980); Cheng *et al., Plant Science Letters, 19*: 91-99 (1980). Regeneration of whole sexually mature *Glycine max* plants from somatic embryos generated from explants of immature *Glycine max* embryos has been reported (Ranch *et al., In Vitro Cellular & Developmental Biology 21*: 653-658 (1985). Regeneration of mature *Glycine max* plants from tissue culture by organogenesis and embryogenesis has also been reported (Barwale *et al., Planta 167*: 473-481 (1986); Wright *et al., Plant Cell Reports 5*: 150-154 (1986)).

The present invention also provides a *Glycine max* plant selected for by screening for an enhanced yield in the *Glycine max* plant, the selection comprising interrogating genomic DNA for the presence of a marker molecule that is genetically linked to an allele of a quantitative trait locus associated with enhanced yield in the *Glycine max* plant, where the allele of a quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

In light of the current disclosure, plant introductions and germplasm can be screened with a marker nucleic acid molecule of the present invention to screen for the presense of a quantitative trait locus associated with enhanced yield in *Glycine max* using one or more of techniques disclosed herein or known in the art.

The present invention also provides a method for screening for enhanced yield comprising interrogating genomic DNA for the presence or absence of a marker molecule that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant between and including U3944117 and within 50 cM of U3944117 or its complement on linkage group U26, where the source of the enhanced yield is *Glycine soja* PI407305 or progeny thereof; and detecting the presence or absence of the marker. Plants having the quantitative trait locus of the present invention may also be selected based on a visible phenotype such as height (*see* Table 4). As used herein, the term "interrogating" refers to any method capable of detecting a feature, such as a polymorphism, of genomic DNA.

As used herein, an agent, be it a naturally occurring molecule or otherwise may be "substantially purified", if desired, referring to a molecule separated from substantially all other molecules normally associated with it in its native state. More preferably a substantially purified molecule is the predominant species present in a preparation. A substantially purified molecule may be greater than 60% free, preferably 75% free, more preferably 90% free, and most preferably 95% free from the other molecules (exclusive of solvent) present in the natural mixture. The term "substantially purified" is not intended to encompass molecules present in their native state.

The agents of the present invention will preferably be "biologically active" with respect to either a structural attribute, such as the capacity of a nucleic acid to hybridize to another nucleic acid molecule, or the ability of a protein to be bound by an antibody (or to compete with another molecule for such binding). Alternatively, such an attribute may be catalytic, and thus involve the capacity of the agent to mediate a chemical reaction or response.

The agents of the present invention may also be recombinant. As used herein, the term recombinant means any agent (*e.g.* DNA, peptide etc.), that is, or results, however indirect, from human manipulation of a nucleic acid molecule.

The agents of the present invention may be labeled with reagents that facilitate detection of the agent (*e.g.* fluorescent labels (Prober *et al.*, *Science* 238:336-340 (1987); Albarella *et al.*, European Patent 144914), chemical labels (Sheldon *et al.*, U.S. Patent 4,582,789; Albarella *et al.*, U.S. Patent 4,563,417), modified bases (Miyoshi *et al.*, European Patent 119448)).

It is further understood, that the present invention provides bacterial, viral, microbial, insect, mammalian and plant cells comprising the agents of the present invention.

Nucleic acid molecules or fragments thereof are capable of specifically hybridizing to other nucleic acid molecules under certain circumstances. As used herein, two nucleic acid molecules are said to be capable of specifically hybridizing to one another if the two molecules are capable of forming an anti-parallel, double-stranded nucleic acid structure. A nucleic acid molecule is said to be the "complement" of another nucleic acid molecule if they exhibit complete complementarity. As used herein, molecules are said to exhibit "complete complementarity" when every nucleotide of one of the molecules is complementary to a nucleotide of the other. Two molecules are said to be "minimally complementary" if they can hybridize to one another with sufficient stability to permit them to remain annealed to one another under at least conventional "low-stringency" conditions. Similarly, the molecules are said to be "complementary" if they can hybridize to one another with sufficient stability to permit them to remain annealed to one another under conventional "high-stringency" conditions. Conventional stringency conditions are described by Sambrook *et al.*, In: *Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press*, Cold Spring Harbor, New York (1989)), and by Haymes *et al.*, In: *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985). Departures from complete complementarity are therefore permissible, as long as such departures do not completely preclude the capacity of the molecules to form a double-stranded structure. In order for a nucleic acid molecule to serve as a primer or probe it need only be sufficiently complementary in sequence to be able to form a stable double-stranded structure under the particular solvent and salt concentrations employed.

As used herein, a substantially homologous sequence is a nucleic acid sequence that will specifically hybridize to the complement of the nucleic acid sequence to which it is being compared under high stringency conditions.

Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low

stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Both temperature and salt may be varied, or either the temperature or the salt concentration may be held constant while the other variable is changed.

In a preferred embodiment, a nucleic acid of the present invention will specifically hybridize to one or more of the nucleic acid molecules set forth in SEQ ID NO: 1 through SEQ ID NO:40 or complements thereof or fragments of either under moderately stringent conditions, for example at about 2.0 x SSC and about 65°C. In a particularly preferred embodiment, a nucleic acid of the present invention will specifically hybridize to one or more of the nucleic acid molecules set forth in SEQ ID NO:1 through SEQ ID NO:40 or complements or fragments of either under high stringency conditions. In one aspect of the present invention, a preferred marker nucleic acid molecule of the present invention has the nucleic acid sequence set forth in SEQ ID NO:1 through SEQ ID NO:40 or complements thereof or fragments of either. In another aspect of the present invention, a preferred marker nucleic acid molecule of the present invention shares between 80% and 100% or 90% and 100% sequence identity with the nucleic acid sequence set forth in SEQ ID NO:1 through SEQ ID NO:40 or complement thereof or fragments of either. In a further aspect of the present invention, a preferred marker nucleic acid molecule of the present invention shares between 95% and 100% sequence identity with the sequence set forth in SEQ ID NO:1 through SEQ ID NO:40 or complement thereof or fragments of either. In a more preferred aspect of the present invention, a preferred marker nucleic acid molecule of the present invention shares between 98% and 100% sequence identity with the nucleic acid sequence set forth in SEQ ID NO:1 through SEQ ID NO:40 or complement thereof or fragments of either.

Additional genetic markers can be used to select plants with an allele of a quantitative trait locus associated with enhanced yield in *Glycine max* of the present invention. In light of the present disclosure, other markers which map to within 50 or less centimorgans of an allele of a quantitative trait locus associated with enhanced yield in *Glycine max* of the present invention will be apparent to those of ordinary skill in the art. Examples of public marker databases include, for example: Soybase, an Agricultural Research Service, United States Department of Agriculture (http://129.186.26.940/ and USDA - Agricultural Research Service:

http://www.ars.usda.gov/). In an embodiment, a genetic marker of the present invention will specifically hybridize in a region between and including marker U3944117 and within 50 cM of U3944117 or its complement on linkage group U26 under moderate stringency. In a preferred embodiment, a genetic marker of the present invention will specifically hybridize in a region between and including marker U3944117 and within 50 cM of U3944117 or its complement on linkage group U26 under high stringency. In a preferred embodiment, a genetic marker of the present invention will specifically hybridize in a region between and including Satt168 and Satt 560 on linkage group U26. In a more preferred embodiment, a genetic marker of the present invention will specifically hybridize in a region between and including Satt168 and Satt 560 on linkage group U26 under high stringency.

A preferred group of markers is selected from the group consisting of a marker nucleic acid molecule that specifically hybridizes to U3944117 (5' GAC TGC GTA CCA ATT CAG AAG TTA TAA GTT GTC ATA AAT ATG AAT CAG TTT CAC TCT GTG ACA ATG ATG GTT CCC TGG ATT-3' (SEQ ID NO: 1) and 5'-GAC TGC GTA CCA ATT CAG AAG TTA TAA GTT GTC ATA AAT ATG CAT CAG TTT CAC TCT ATG ATA ATG ATG GTT CCA TGG ATT-3' (SEQ ID NO: 2) (U3944117b sequence) or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt168 (5'-CGC TTG-CCC AAA AAT TAA TAG TAT GAT CAT ATG TGG TTG GAG GGT GGG AAA AAA AAG AAG ACA ATA CCA TAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAG AGA AAA TAC ATC AAT ACT AAG AAG TTA TTA ATT TAA ATG ATA CTG AAT TTA ATA TCC TTA ANT TAA TTC TCC NAA AGA NAT ATA AGA TTG AGG TTG GAG AAT GG-3' (SEQ ID NO:3) or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt416 (5-'TAT AGC CCA GCA AAA AAA AAC AGA -GAT TAA GAG AAG ACG AGA GTT TTA AGA GAA TAA GAA AAA TTT TGT GTA TTC TTT CTA AGA AGA ATA AAT TAT TTA TAA ATA CAA AAT ATA ATT GAA AAA ATA TAA AAA ATA AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAC AAA CAA TCA TAA AAG ATA ATT AAA GAT ATG AAC AAT CAC ACA GAT AAA TTA CCC ATA ACA AAT ATT ACT AAA ATA CTA AAA TTA TGT TAT TAA TAT AAC CGA TTT TTT TGT TCA TTG GTC GGT TTT GAT-3') (SEQ ID NO:4) or its complement, a marker nucleic acid molecule that specifically

hybridizes to Sat_083 (5'-ACC ATT GGA ATG TTC TAC AAT TAA GAA TAA AAT CTT TAA CAT TAG GAA AAA AAT ATA AAA AAT ATT AAA TAA ATA TTA AAT AAA ACT AAA ACT AAA AAT ATT AAT TTA AAT AAA ATT TTT TAC TAT TTT AAA GTG GGA ATA TAT ATA TAT ATA TAT ATA TAT ATA TAT ATA TAT ATA TAT ATA TAT NNN TTT GGG ACT TCA AAA TCA TAT TTT AAA AAA ATT AAG AAG ATG TAA ACT TTT TTA TAA CTT CAA-3') (SEQ ID NO:5) or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt474 (5'-AAT TTG GAA ATG ACA TCT TAG AAA TTA TTC TCA CAA CTC TTG TAA TAT ACA AGT ATT AAG AAT GAT TTT ACT AAC TTA ATA AAA TTC AGA AAA AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAA AGA TTG GGA AGG CAG ATA GAA GAA TAT ATA TGT TCT CAC CGC AAT ACT TGG TCG TTT TGT AAT ATT TGT AGC CCA ACA TAT AGC AGT ATC TCT TTT CTT CAC ATC CAA TTT CTC CCG T-3') (SEQ ID NO:6) or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt122 (5'-AAC CAA CTT GGG AAT AGA CAA TAA TTC AAG AAA TAC AAG TGC AAG AAA GAC CTA ATA ATA ATA ATA ATA ATA ATA ATC CTA AAA ATG GAG TTA ATG TCT TGG TAT GAT TAG TGA ATG ATA GAG AGC -3') (SEQ ID NO:7) or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt556 (5'-ATA AAA CCC GAT AAA TAA GAT TTC ATG AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT ACT CAA AGA CCA AAA TTT CAT TTT CGA AAC ATG ATA TAG GCT TCA GAG ATG AAC GAA CAT AAA ATA CAT AAG AAA ACA AGG TGC ACA-3') (SEQ ID NO:8) or its complement, a marker nucleic acid molecule that specifically hybridizes to Sct_094 (5'-GGG TGA AGT GAG AGT AAC ACG TAA GAG TNC CTC TCT CTC TCT CTC TCT CTC TCT CTC TCT CTC TCT CTC TCT AAT ACT AGG GGG AAG TTA TGT CTA CCA ATG AAG AGA TCC GGG -3') (SEQ ID NO:9) or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt272 (5'-ATG ACA AGG AAA AAT CAA TCA ACA ATC TGA ACC TTT TTC CAC TTT TTT CCT TGT TCA ACA TTA TAA GGT TGC TCA CAT ATT ATA TAA AGA TTT CAT GTT CTC TCA CTC ACC TCA TAC ACC ATT CCT ACT CTC TTA AAC ACA CAC ATA CTT TAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAA CCA TTC CA AAC ACT CTT AAC AGC AGC -3') (SEQ ID NO:10) or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt020 (5'-GAG AAA GAA

ATG TGT TAG TGT AAT AAA AAA GAC TAA AAT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT TGT GCA ATC AAA CAA TAA GAA GGA AAA G-3') (SEQ ID NO:11) or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt066 (5'-TNN NCN ACG CCG CTT GAT AAA AAC ACA AAT TTA TAA TAA TCA AAA ACA TAT TTA AGC TTA ATA ATG AAA ATG ACA CCA TTA AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT ANT AAT AAT AAT AAT AAT AAT AAT AAT AAT CAC AAC AAA AAT AGN TCA TGT AAA ATG GAA TGT TAC AGA AGT GAT CAA-3') (SEQ ID NO:12) or its complement, a marker nucleic acid molecule that specifically hybridizes to Satt534 (5'-CTC CTC CTG CGC AAC AAC AAT ATT CAT GCA TAT ACA TCA CGT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATT ATA AAA TGC AGG TAA GAC ATT CAA CAA GAT AAC TAA GGT CAT GGC CTA GAT CCC CC-3') (SEQ ID NO:13) or its complement, and a marker nucleic acid molecule that specifically hybridizes to Satt560 (5'-ATC GTG CAA GAA AAT AAA TTT TTG AAA ATA ATT AAT ATT TTA ATT TTT AAT ATG ATA TTA ATT ATT TTT TAT TTA TAT TTT TTA TAT ATT ACT AGT AAA CAA AAT TTA AAA ATA AAT TAA TAA TAA TAA AAT ATT AAT TTT ATC TAT ATA TTA TTA TTA TTA TTA TTA TTA TTA TTA TTA TTA TTA TTA TCA TTA TTA TTA TTA ATA GTG TGC AAA ACA AGT TAT TGT AAT AAG ATA ATT ATT TAG AGA CGG ATG AAG TAA TTA TTT GAG GCG AAG TCC AC -3') (SEQ ID NO:14) or its complement. In a preferred embodiment, the genetic marker of the present invention is an SSR or AFLP marker.

Polymorphisms may also be found using a DNA fingerprinting technique called amplified fragment length polymorphism (AFLP), which is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA to profile that DNA (Vos *et al.*, *Nucleic Acids Res. 23*:4407-4414 (1995)). This method allows for the specific co-amplification of high numbers of restriction fragments, which can be visualized by PCR without knowledge of the nucleic acid sequence.

AFLP employs basically three steps. Initially, a sample of genomic DNA is cut with restriction enzymes and oligonucleotide adapters are ligated to the restriction fragments of the DNA. The restriction fragments are then amplified using PCR by

using the adapter and restriction sequence as target sites for primer annealing. The selective amplification is achieved by the use of primers that extend into the restriction fragments, amplifying only those fragments in which the primer extensions match the nucleotide flanking the restriction sites. These amplified fragments are then visualized on a denaturing polyacrylamide gel.

AFLP analysis has been performed on *Salix* (Beismann *et al.*, *Mol. Ecol.* *6*:989-993 (1997)), *Acinetobacter* (Janssen *et al.*, *Int. J. Syst. Bacteriol. 47*:1179-1187 (1997)), *Aeromonas popoffi* (Huys *et al.*, *Int. J. Syst. Bacteriol. 47*:1165-1171 (1997)), rice (McCouch *et al.*, *Plant Mol. Biol. 35*:89-99 (1997); Nandi *et al.*, *Mol. Gen. Genet.* *255*:1-8 (1997); Cho *et al.*, *Genome 39*:373-378 (1996)), barley (*Hordeum vulgare*) (Simons *et al.*, *Genomics 44*:61-70 (1997); Waugh *et al.*, *Mol. Gen. Genet. 255*:311-321 (1997); Qi *et al.*, *Mol. Gen Genet. 254*:330-336 (1997); Becker *et al.*, *Mol. Gen. Genet. 249*:65-73 (1995)), potato (Van der Voort *et al.*, *Mol. Gen. Genet. 255*:438-447 (1997); Meksem *et al.*, *Mol. Gen. Genet. 249*:74-81 (1995)), *Phytophthora infestans* (Van der Lee *et al.*, *Fungal Genet. Biol. 21*:278-291 (1997)), *Bacillus anthracis* (Keim *et al.*, *J. Bacteriol. 179*:818-824 (1997)), *Astragalus cremnophylax* (Travis *et al.*, *Mol. Ecol. 5*:735-745 (1996)), *Arabidopsis thaliana* (Cnops *et al.*, *Mol. Gen. Genet. 253*:32-41 (1996)), *Escherichia coli* (Lin *et al.*, *Nucleic Acids Res. 24*:3649-3650 (1996)), *Aeromonas* (Huys *et al.*, *Int. J. Syst. Bacteriol. 46*:572-580 (1996)), nematode (Folkertsma *et al.*, *Mol. Plant Microbe Interact. 9*:47-54 (1996)), tomato (Thomas *et al.*, *Plant J. 8*:785-794 (1995)), and human (Latorra *et al.*, *PCR Methods Appl. 3*:351-358 (1994)). AFLP analysis has also been used for fingerprinting mRNA (Money *et al.*, *Nucleic Acids Res. 24*:2616-2617 (1996),; Bachem *et al.*, *Plant J.* *9*:745-753 (1996)). It is understood that one or more of the nucleic acids of the present invention, can be utilized as markers or probes to detect polymorphisms by AFLP analysis or for fingerprinting RNA.

In a preferred embodiment, a marker molecule is detected by DNA amplification using a forward and a reverse primer capable of detecting a marker molecule of the present invention. In a particularly preferred embodiment, a marker molecule is detected by AFLP amplification.

Microsatellite (SSR) markers have been used to distinguish the genotype of soybean cultivars and elite breeding lines. These methods have been developed for soybean and are well known in the field of molecular plant breeding (Rongwen, *Theor. Appl. Gen. 90*:43-48 (1995); Akkaya, *Crop Sci. 35*:1439-1445 (1995); Mansur,

Crop Sci. 36:1327-1336 (1996); Diwan, *Theor. Appl. Gen. 95*:723-733 (1997); Simple sequence repeat DNA marker analysis, in "DNA markers: Protocols, applications, and overviews: (1997) 173-185, Cregan, et al., eds., Wiley-Liss NY. In a particularly preferred embodiment, a marker molecule is detected by SSR techniques. It is understood that SSR and AFLP primers can hybridize to a combination of plant DNA and adapter DNA (*e.g. Eco*RI adapter or *MseI* adapter, Vos *et al.*, *Nucleic Acids Res. 23*:4407-4414 (1995)). In a particularly preferred embodiment, U3944117 can be detected by using a forward primer and a reverse primer, the forward primer having the nucleic acid sequence 5'-GAC TGC GTA CCA ATT C AGA-3' (E39 primer) (SEQ ID NO:15) and the reverse primer having the nucleic acid sequence 5-GAT GAG TCC TGA GTA A ATC-3' (M44 primer) (SEQ ID NO:16); Satt168 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'- CGCTTGCCCAAAAATTAATAGTA-3' (SEQ ID NO:17) and the reverse primer having the sequence 5'- CCA TTC TCC AAC CTC AAT CTT ATA T -3' (SEQ ID NO:18); Satt416 can be detected using a forward and a reverse primer, the forward primer having the sequence 5'- TAT AGC CCA GCA AAA AAA AAC AGA GAT-3' (SEQ ID NO:19) and the reverse primer having the sequence 5'- ATC AAA ACC GAC CAA TGA ACA AAA AAA-3' (SEQ ID NO:20); Sat_083 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'-ACC ATT GGA ATG TTC TAC A -3' (SEQ ID NO:21) and the reverse primer having the sequence 5'-TTG AAG TTA TAA AAA AGT TTA CAT C -3' (SEQ ID NO:22); Satt474 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'-GCG AAA TTT GGA AAT GAC ATC TTA GAA -3' (SEQ ID NO:23) and the reverse primer having the sequence 5'-GCG ACG GGA GAA ATT GGA TGT GAA GAA -3' (SEQ ID NO:24); Satt122 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'-AAC CAA CTT GGG AAT AGA C -3' (SEQ ID NO:25) and the reverse primer having the sequence 5'-GCT CTC TAT CAT TCA CTA ATC A -3' (SEQ ID NO:26); Satt556 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'-GCG ATA AAA CCC GAT AAA TAA -3' (SEQ ID NO:27) and the reverse primer having the sequence 5-GCG TTG TGC ACC TTG TTT TCT-3' (SEQ ID NO:28); and Sct_094 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'-GGG TGA AGT GAG AGT AAC A-3' (SEQ ID NO:29) and the reverse primer having the

sequence 5'-CCC GGA TCT CTT CAT T -3' (SEQ ID NO:30); and Satt272 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'-ATG ACA AGG AAA AAT CAA TCA AC -3' (SEQ ID NO:31) and the reverse primer having the sequence 5'-GCT GCT GTT AAG AGT GTT TG -3' (SEQ ID NO:32); and Satt020 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'-TTT GAA GGA AGG GTG GTG AG-3' (SEQ ID NO:33) and the reverse primer having the sequence 5' GAT CCA AAT CCT CAG TAT CAT A-3' (SEQ ID NO:34); and Satt066 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'-GGG AAG CTT AAT AAT GAA AAT GAC AC-3' (SEQ ID NO:35) and the reverse primer having the sequence 5'-TTG ATC ACT TCT GTA ACA TTC -3' (SEQ ID NO:36); and Satt534 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'-CTC CTC CTG CGC AAC AAC AAT A -3' (SEQ ID NO:37) and the reverse primer having the sequence 5'-GGG GGA TCT AGG CCA TGA C -3' (SEQ ID NO:38); and Satt560 can be detected using a forward primer and a reverse primer, the forward primer having the sequence 5'-GCG ATC GTG CAA GAA AAT A -3' (SEQ ID NO:39) and the reverse primer having the sequence 5'-GCG GTG GAC TTC GCC TCA AAT AAT -3' (SEQ ID NO:40). Other primers that recognize other adapter sequences can be used.

Genetic markers of the present invention include "dominant" or "codominant" markers. "Codominant markers" reveal the presence of two or more alleles (two per diploid individual). "Dominant markers" reveal the presence of only a single allele. The presence of the dominant marker phenotype (*e.g.*, a band of DNA) is an indication that one allele is present in either the homozygous or heterozygous condition. The absence of the dominant marker phenotype (*e.g.*, absence of a DNA band) is merely evidence that "some other" undefined allele is present. In the case of populations where individuals are predominantly homozygous and loci are predominantly dimorphic, dominant and codominant markers can be equally valuable. As populations become more heterozygous and multiallelic, codominant markers often become more informative of the genotype than dominant markers.

Additional markers, such as microsatellite markers (SSR), AFLP markers, RFLP markers, RAPD markers, phenotypic markers, SNPs, isozyme markers, microarray transcription profiles that are genetically linked to or correlated with alleles of a QTL of the present invention can be utilized (Walton, *Seed World* 22-29

(July, 1993); Burow and Blake, *Molecular Dissection of Complex Traits,* 13-29, Eds. Paterson, CRC Press, New York (1988)). Methods to isolate such markers are known in the art. For example, locus-specific microsatellite markers (SSR) can be obtained by screening a genomic library for microsatellite repeats, sequencing of "positive" clones, designing primers which flank the repeats, and amplifying genomic DNA with these primers. The size of the resulting amplification products can vary by integral numbers of the basic repeat unit. To detect a polymorphism, PCR products can be radiolabeled, separated on denaturing polyacrylamide gels, and detected by autoradiography. Fragments with size differences >4 bp can also be resolved on agarose gels, thus avoiding radioactivity.

Other microsatellite markers may be utilized. Amplification of simple tandem repeats, mainly of the $[CA]_n$ type were reported by Litt and Luty, *Amer. J. Human Genet. 44*:397-401 (1989); Smeets *et al., Human Genet. 83*:245-251 (1989); Tautz, *Nucleic Acids Res. 17*:6463-6472 (1989); Weber and May, *Am. J. Hum. Genet. 44*:388-396 (1989). Weber, *Genomics 7*:524-530 (1990), reported that the level of polymorphism detected by PCR-amplified $[CA]_n$ type microsatellites depends on the number of the "perfect" (*i.e.*, uninterrupted), tandemly repeated motifs. Below a certain threshold (*i.e.*, 12 CA-repeats), the microsatellites were reported to be primarily monomorphic. Above this threshold, however, the probability of polymorphism increases with microsatellite length. Consequently, long, perfect arrays of microsatellites are preferred for the generation of markers, *i.e.*, for the design and synthesis of flanking primers.

Suitable primers can be deduced from DNA databases (*e.g.*, Akkaya *et al., Genetics. 132:*1131-1139 (1992)). Alternatively, size-selected genomic libraries (200 to 500 bp) can be constructed by, for example, using the following steps: (1) isolation of genomic DNA; (2) digestion with one or more 4 base-specific restriction enzymes; (3) size-selection of restriction fragments by agarose gel electrophoresis, excision and purification of the desire size fraction; (4) ligation of the DNA into a suitable vector and transformation into a suitable *E. coli* strain; (5) screening for the presence of microsatellites by colony or plaque hybridization with a labeled probe; (6) isolation of positive clones and sequencing of the inserts; and (7) design of suitable primers flanking the microsatellite repeat.

Establishing libraries with small, size-selected inserts can be advantageous for microsatellite isolation for two reasons: (1) long microsatellites are often unstable in

E. coli, and (2) positive clones can be sequenced without subcloning. A number of approaches have been reported for the enrichment of microsatellites in genomic libraries. Such enrichment procedures are particularly useful if libraries are screened with comparatively rare tri- and tetranucleotide repeat motifs. One such approach has been described by Ostrander *et al.*, *Proc. Natl. Acad. Sci. (U.S.A). 89*:3419-3423 (1992), who reported the generation of a small-insert phagemid library in an *E. coli* strain deficient in UTPase (d8t) and uracil-N-glycosylase (ung) genes. In the absence of UTPase and uracil-N-glycosylase, dUTP can compete with dTTP for the incorporation into DNA. Single-stranded phagemid DNA isolated from such a library, can be primed with $[CA]_n$ and $[TG]_n$ primers for second strand synthesis, and the products used to transform a wild-type *E. coli* strain. Since under these conditions there will be selection against single-stranded, uracil-containing DNA molecules, the resulting library will consist of primer-extended, double-stranded products and an about 50-fold enrichment in CA-repeats.

Other reported enrichment strategies rely on hybridization selection of simple sequence repeats prior to cloning (Karagyozov *et al.*, *Nucleic Acids Res. 21*:3911-3912 (1993); Armour *et al.*, *Hum. Mol. Gen. 3*:599-605 (1994); Kijas *et al.*, *Genome 38*:349-355 (1994); Kandpal *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 91*:88-92 (1994); Edwards *et al.*, *Am. J. Hum. Genet. 49*:746-756 (1991)). Hybridization selection, can for example, involve the following steps: (1) genomic DNA is fragmented, either by sonication, or by digestion with a restriction enzyme; (2) genomic DNA fragments are ligated to adapters that allow a "whole genome PCR" at this or a later stage of the procedure; (3) genomic DNA fragments are amplified, denatured and hybridized with single-stranded microsatellite sequences bound to a nylon membrane; (4) after washing off unbound DNA, hybridizing fragments enriched for microsatellites are eluted from the membrane by boiling or alkali treatment, reamplified using adapter-complementary primers, and digested with a restriction enzyme to remove the adapters; and (5) DNA fragments are ligated into a suitable vector and transformed into a suitable *E. coli* strain. Microsatellite can be found in up to 50-70% of the clones obtained from these procedures (Armour *et al.*, *Hum. Mol. Gen. 3*:599-605 (1994); Edwards *et al.*, *Am. J. Hum. Genet. 49*:746-756 (1991).

An alternative hybridization selection strategy was reported by Kijas *et al.*, *Genome 38*:599-605 (1994), which replaced the nylon membrane with biotinylated, microsatellite-complementary oligonucleotides attached to streptavidin-coated

magnetic particles. Microsatellite-containing DNA fragments are selectively bound to the magnetic beads, reamplified, restriction-digested and cloned.

It is further understood that other additional markers on linkage group U26 (B2) may be utilized (Morgante *et al., Genome 37:*763-769 (1994)). PCR-amplified microsatellites can be used, because they are locus-specific, codominant, occur in large numbers and allow the unambiguous identification of alleles. Standard PCR-amplified microsatellites protocols use radioisotopes and denaturing polyacrylamide gels to detect amplified microsatellites. In many situations, however, allele sizes are sufficiently different to be resolved on high percentage agarose gels in combination with ethidium bromide staining (Bell and Ecker, *Genomics 19:*137-144 (1994); Becker and Heun, *Genome 38*:991-998 (1995); Huttel, Ph.D. Thesis, University of Frankfurt, Germany (1996)). High resolution without applying radioactivity is also provided by nondenaturing polyacrylamide gels in combination with either ethidium bromide (Scrimshaw, *Biotechniques 13*:2189 (1992)) or silver straining (Klinkicht and Tautz, *Molecular Ecology 1*: 133-134 (1992); Neilan *et al., Biotechniques 17*:708-712 (1994)). An alternative of PCR-amplified microsatelllites typing involves the use of fluorescent primers in combination with a semi-automated DNA sequencer (Schwengel *et al., Genomics 22*:46-54 (1994)). Fluorescent PCR products can be detected by real-time laser scanning during gel electrophoresis. An advantage of this technology is that different amplification reactions as well as a size marker (each labeled with a different fluorophore) can be combined into one lane during electrophoresis. Multiplex analysis of up to 24 different microsatellite loci per lane has been reported (Schwengel *et al., Genomics 22*:46-54 (1994)).

The detection of polymorphic sites in a sample of DNA may be facilitated through the use of nucleic acid amplification methods. Such methods specifically increase the concentration of polynucleotides that span the polymorphic site, or include that site and sequences located either distal or proximal to it. Such amplified molecules can be readily detected by gel electrophoresis or other means.

The most preferred method of achieving such amplification employs the polymerase chain reaction ("PCR") (Mullis *et al., Cold Spring Harbor Symp. Quant. Biol. 51*:263-273 (1986); Erlich *et al.*, European Patent Appln. 50,424; European Patent Appln. 84,796, European Patent Application 258,017, European Patent Appln. 237,362; Mullis, European Patent Appln. 201,184; Mullis *et al.*, U.S. Patent No. 4,683,202; Erlich, U.S. Patent No. 4,582,788; and Saiki *et al.*, U.S. Patent No.

4,683,194), using primer pairs that are capable of hybridizing to the proximal sequences that define a polymorphism in its double-stranded form.

In lieu of PCR, alternative methods, such as the "Ligase Chain Reaction" ("LCR") may be used (Barany, *Proc. Natl. Acad. Sci. (U.S.A.) 88*:189-193 (1991)). LCR uses two pairs of oligonucleotide probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependent ligase. As with PCR, the resulting products thus serve as a template in subsequent cycles and an exponential amplification of the desired sequence is obtained.

LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a polymorphic site. In one embodiment, either oligonucleotide will be designed to include the actual polymorphic site of the polymorphism. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the polymorphic site present on the oligonucleotide. Alternatively, the oligonucleotides may be selected such that they do not include the polymorphic site (see, Segev, PCT Application WO 90/01069).

The "Oligonucleotide Ligation Assay" ("OLA") may alternatively be employed (Landegren *et al.*, *Science 241*:1077-1080 (1988)). The OLA protocol uses two oligonucleotides that are designed to be capable of hybridizing to abutting sequences of a single strand of a target. OLA, like LCR, is particularly suited for the detection of point mutations. Unlike LCR, however, OLA results in "linear" rather than exponential amplification of the target sequence.

Nickerson *et al.* have described a nucleic acid detection assay that combines attributes of PCR and OLA (Nickerson *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 87*:8923-8927 (1990)). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA. In addition to requiring multiple, and separate, processing steps, one problem associated with such combinations is that they inherit all of the problems associated with PCR and OLA.

Schemes based on ligation of two (or more) oligonucleotides in the presence of a nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby

amplifying the di-oligonucleotide, are also known (Wu *et al.*, *Genomics 4*:560-569 (1989)), and may be readily adapted to the purposes of the present invention.

Other known nucleic acid amplification procedures, such as allele-specific oligomers, branched DNA technology, transcription-based amplification systems, or isothermal amplification methods may also be used to amplify and analyze such polymorphisms (Malek *et al.*, U.S. Patent 5,130,238; Davey *et al.*, European Patent Application 329,822; Schuster *et al.*, U.S. Patent 5,169,766; Miller *et al.*, PCT Patent Application WO 89/06700; Kwoh, *et al.*, *Proc. Natl. Acad. Sci.* (U.S.A.) *86*:1173-1177 (1989); Gingeras *et al.*, PCT Patent Application WO 88/10315; Walker *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 89*:392-396 (1992)).

Polymorphisms can also be identified by Single Strand Conformation Polymorphism (SSCP) analysis. SSCP is a method capable of identifying most sequence variations in a single strand of DNA, typically between 150 and 250 nucleotides in length (Elles, *Methods in Molecular Medicine: Molecular Diagnosis of Genetic Diseases*, Humana Press (1996); Orita *et al., Genomics 5:* 874-879 (1989)). Under denaturing conditions a single strand of DNA will adopt a conformation that is uniquely dependent on its sequence conformation. This conformation usually will be different, even if only a single base is changed. Most conformations have been reported to alter the physical configuration or size sufficiently to be detectable by electrophoresis. A number of protocols have been described for SSCP including, but not limited to, Lee *et al., Anal. Biochem. 205:* 289-293 (1992); Suzuki *et al., Anal. Biochem. 192:* 82-84 (1991); Lo *et al., Nucleic Acids Research 20:* 1005-1009 (1992); Sarkar *et al., Genomics 13:*441-443 (1992). It is understood that one or more of the nucleic acids of the present invention, can be utilized as markers or probes to detect polymorphisms by SSCP analysis.

Polymorphisms may also be found using random amplified polymorphic DNA (RAPD) (Williams *et al., Nucl. Acids Res. 18*: 6531-6535 (1990)) and cleaveable amplified polymorphic sequences (CAPS) (Lyamichev *et al., Science 260*: 778-783 (1993)). It is understood that one or more of the nucleic acid molecules of the present invention, can be utilized as markers or probes to detect polymorphisms by RAPD or CAPS analysis.

The identification of a polymorphism can be determined in a variety of ways. By correlating the presence or absence of it in a plant with the presence or absence of a phenotype, it is possible to predict the phenotype of that plant. If a polymorphism

creates or destroys a restriction endonuclease cleavage site, or if it results in the loss or insertion of DNA (*e.g.*, a variable nucleotide tandem repeat (VNTR) polymorphism), it will alter the size or profile of the DNA fragments that are generated by digestion with that restriction endonuclease. As such, individuals that possess a variant sequence can be distinguished from those having the original sequence by restriction fragment analysis. Polymorphisms that can be identified in this manner are termed "restriction fragment length polymorphisms" ("RFLPs"). RFLPs have been widely used in human and plant genetic analyses (Glassberg, UK Patent Application 2135774; Skolnick *et al.*, *Cytogen. Cell Genet. 32*:58-67 (1982); Botstein *et al.*, *Ann. J. Hum. Genet. 32*:314-331 (1980); Fischer *et al.* (PCT Application WO90/13668); Uhlen, PCT Application WO90/11369).

A central attribute of "single nucleotide polymorphisms," or "SNPs" is that the site of the polymorphism is at a single nucleotide. SNPs have certain reported advantages over RFLPs and VNTRs. First, SNPs are more stable than other classes of polymorphisms. Their spontaneous mutation rate is approximately 10^{-9} (Kornberg, DNA Replication, W. H. Freeman & Co., San Francisco, 1980), approximately 1,000 times less frequent than VNTRs (U.S. Patent 5,679,524). Second, SNPs occur at greater frequency, and with greater uniformity than RFLPs and VNTRs. As SNPs result from sequence variation, new polymorphisms can be identified by sequencing random genomic or cDNA molecules. SNPs can also result from deletions, point mutations and insertions. Any single base alteration, whatever the cause, can be a SNP. The greater frequency of SNPs means that they can be more readily identified than the other classes of polymorphisms.

SNPs can be characterized using any of a variety of methods. Such methods include the direct or indirect sequencing of the site, the use of restriction enzymes where the respective alleles of the site create or destroy a restriction site, the use of allele-specific hybridization probes, the use of antibodies that are specific for the proteins encoded by the different alleles of the polymorphism or by other biochemical interpretation. SNPs can sequenced by a number of methods. Two basic methods may be used for DNA sequencing, the chain termination method of Sanger *et al., Proc. Natl. Acad. Sci. (U.S.A.) 74:* 5463-5467 (1977), and the chemical degradation method of Maxam and Gilbert, *Proc. Nat. Acad. Sci. (U.S.A.) 74:* 560-564 (1977). Automation and advances in technology such as the replacement of radioisotopes with fluorescence-based sequencing have reduced the effort required to sequence DNA

(Craxton, *Methods, 2:* 20-26 (1991); Ju *et al., Proc. Natl. Acad. Sci. (U.S.A.) 92:* 4347-4351 (1995); Tabor and Richardson, *Proc. Natl. Acad. Sci. (U.S.A.) 92:* 6339-6343 (1995)). Automated sequencers are available from, for example, Pharmacia Biotech, Inc., Piscataway, New Jersey (Pharmacia ALF), LI-COR, Inc., Lincoln, Nebraska (LI-COR 4,000) and Millipore, Bedford, Massachusetts (Millipore BaseStation).

In addition, advances in capillary gel electrophoresis have also reduced the effort required to sequence DNA and such advances provide a rapid high resolution approach for sequencing DNA samples (Swerdlow and Gesteland, *Nucleic Acids Res. 18*:1415-1419 (1990); Smith, *Nature 349*:812-813 (1991); Luckey *et al., Methods Enzymol. 218*:154-172 (1993); Lu *et al., J. Chromatog. A. 680*:497-501 (1994); Carson *et al., Anal. Chem. 65*:3219-3226 (1993); Huang *et al., Anal. Chem. 64*:2149-2154 (1992); Kheterpal *et al., Electrophoresis 17*:1852-1859 (1996); Quesada and Zhang, *Electrophoresis 17*:1841-1851 (1996); Baba, *Yakugaku Zasshi 117*:265-281 (1997), Marino, Appl. Theor. Electrophor. 5:1-5 (1995)).

A microarray-based method for high-throughput monitoring of plant gene expression can be utilized as a genetic marker system. This 'chip'-based approach involves using microarrays of nucleic acid molecules as gene-specific hybridization targets to quantitatively or qualitatively measure expression of plant genes (Schena *et al., Science 270*:467-470 (1995); Shalon, Ph.D. Thesis. Stanford University (1996)). Every nucleotide in a large sequence can be queried at the same time. Hybridization can be used to efficiently analyze nucleotide sequences. Such microarrays can be probed with any combination of nucleic acid molecules. Particularly preferred combinations of nucleic acid molecules to be used as probes include a population of mRNA molecules from a known tissue type or a known developmental stage or a plant subject to a known stress (environmental or man-made) or any combination thereof (*e.g.* mRNA made from water stressed leaves at the 2 leaf stage). Expression profiles generated by this method can be utilized as markers.

The genetic linkage of additional marker molecules can be established by a gene mapping model such as, without limitation, the flanking marker model reported by Lander and Botstein, *Genetics, 121*:185-199 (1989), and the interval mapping, based on maximum likelihood methods described by Lander and Botstein, *Genetics, 121*:185-199 (1989), and implemented in the software package MAPMAKER/QTL (Lincoln and Lander, *Mapping Genes Controlling Quantitative Traits Using*

MAPMAKER/QTL, Whitehead Institute for Biomedical Research, Massachusetts, (1990). Additional software includes Qgene, Version 2.23 (1996), Department of Plant Breeding and Biometry, 266 Emerson Hall, Cornell University, Ithaca, NY). Use of Qgene software is a particularly preferred approach.

A maximum likelihood estimate (MLE) for the presence of a marker is calculated, together with an MLE assuming no QTL effect, to avoid false positives. A $\log_{10}$ of an odds ratio (LOD) is then calculated as: LOD = $\log_{10}$ (MLE for the presence of a QTL/MLE given no linked QTL).

The LOD score essentially indicates how much more likely the data are to have arisen assuming the presence of a QTL than in its absence. The LOD threshold value for avoiding a false positive with a given confidence, say 95%, depends on the number of markers and the length of the genome. Graphs indicating LOD thresholds are set forth in Lander and Botstein, *Genetics, 121*:185-199 (1989), and further described by Arús and Moreno-González , *Plant Breeding*, Hayward, Bosemark, Romagosa (eds.) Chapman & Hall, London, pp. 314-331 (1993).

Additional models can be used. Many modifications and alternative approaches to interval mapping have been reported, including the use non-parametric methods (Kruglyak and Lander, *Genetics, 139:*1421-1428 (1995)). Multiple regression methods or models can be also be used, in which the trait is regressed on a large number of markers (Jansen, *Biometrics in Plant Breed*, van Oijen, Jansen (eds.) Proceedings of the Ninth Meeting of the Eucarpia Section Biometrics in Plant Breeding, The Netherlands, pp. 116-124 (1994); Weber and Wricke, *Advances in Plant Breeding*, Blackwell, Berlin, 16 (1994)). Procedures combining interval mapping with regression analysis, whereby the phenotype is regressed onto a single putative QTL at a given marker interval, and at the same time onto a number of markers that serve as 'cofactors,' have been reported by Jansen and Stam, *Genetics, 136*:1447-1455 (1994) and Zeng, *Genetics, 136*:1457-1468 (1994). Generally, the use of cofactors reduces the bias and sampling error of the estimated QTL positions (Utz and Melchinger, *Biometrics in Plant Breeding*, van Oijen, Jansen (eds.) Proceedings of the Ninth Meeting of the Eucarpia Section Biometrics in Plant Breeding, The Netherlands, pp.195-204 (1994), thereby improving the precision and efficiency of QTL mapping (Zeng, *Genetics, 136*:1457-1468 (1994)). These models can be extended to multi-environment experiments to analyze genotype-environment interactions (Jansen *et al., Theo. Appl. Genet. 91:*33-37 (1995)).

Selection of an appropriate mapping populations is important to map construction. The choice of appropriate mapping population depends on the type of marker systems employed (Tanksley *et al., Molecular mapping plant chromosomes. chromosome structure and function: Impact of new concepts* J.P. Gustafson and R. Appels (eds.). Plenum Press, New York, pp. 157-173 (1988)). Consideration must be given to the source of parents (adapted vs. exotic) used in the mapping population. Chromosome pairing and recombination rates can be severely disturbed (suppressed) in wide crosses (adapted x exotic) and generally yield greatly reduced linkage distances. Wide crosses will usually provide segregating populations with a relatively large array of polymorphisms when compared to progeny in a narrow cross (adapted x adapted).

An F_2 population is the first generation of selfing after the hybrid seed is produced. Usually a single F_1 plant is selfed to generate a population segregating for all the genes in Mendelian (1:2:1) fashion. Maximum genetic information is obtained from a completely classified F_2 population using a codominant marker system (Mather, Measurement of Linkage in Heredity: Methuen and Co., (1938)). In the case of dominant markers, progeny tests (e.g F_3, BCF_2) are required to identify the heterozygotes, thus making it equivalent to a completely classified F_2 population. However, this procedure is often prohibitive because of the cost and time involved in progeny testing. Progeny testing of F_2 individuals is often used in map construction where phenotypes do not consistently reflect genotype (*e.g.* disease resistance) or where trait expression is controlled by a QTL. Segregation data from progeny test populations (*e.g.* F_3 or BCF_2) can be used in map construction. Marker-assisted selection can then be applied to cross progeny based on marker-trait map associations (F_2, F_3), where linkage groups have not been completely disassociated by recombination events (*i.e.*, maximum disequilibrium).

Recombinant inbred lines (RIL) (genetically related lines; usually $>F_5$, developed from continuously selfing F_2 lines towards homozygosity) can be used as a mapping population. Information obtained from dominant markers can be maximized by using RIL because all loci are homozygous or nearly so. Under conditions of tight linkage (*i.e.*, about <10% recombination), dominant and co-dominant markers evaluated in RIL populations provide more information per individual than either marker type in backcross populations (Reiter *et al., Proc. Natl. Acad. Sci. (U.S.A.) 89*:1477-1481 (1992)). However, as the distance between markers becomes larger

(*i.e.*, loci become more independent), the information in RIL populations decreases dramatically when compared to codominant markers.

Backcross populations (*e.g.*, generated from a cross between a successful variety (recurrent parent) and another variety (donor parent) carrying a trait not present in the former) can be utilized as a mapping population. A series of backcrosses to the recurrent parent can be made to recover most of its desirable traits. Thus a population is created consisting of individuals nearly like the recurrent parent but each individual carries varying amounts or mosaic of genomic regions from the donor parent. Backcross populations can be useful for mapping dominant markers if all loci in the recurrent parent are homozygous and the donor and recurrent parent have contrasting polymorphic marker alleles (Reiter *et al., Proc. Natl. Acad. Sci. (U.S.A.) 89*:1477-1481 (1992)). Information obtained from backcross populations using either codominant or dominant markers is less than that obtained from F_2 populations because one, rather than two, recombinant gametes are sampled per plant. Backcross populations, however, are more informative (at low marker saturation) when compared to RILs as the distance between linked loci increases in RIL populations (*i.e.* about .15% recombination). Increased recombination can be beneficial for resolution of tight linkages, but may be undesirable in the construction of maps with low marker saturation.

Near-isogenic lines (NIL) created by many backcrosses to produce an array of individuals that are nearly identical in genetic composition except for the trait or genomic region under interrogation can be used as a mapping population. In mapping with NILs, only a portion of the polymorphic loci are expected to map to a selected region.

Bulk segregant analysis (BSA) is a method developed for the rapid identification of linkage between markers and traits of interest (Michelmore, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 88*:9828-9832 (1991)). In BSA, two bulked DNA samples are drawn from a segregating population originating from a single cross. These bulks contain individuals that are identical for a particular trait (resistant or susceptible to particular disease) or genomic region but arbitrary at unlinked regions (*i.e.* heterozygous). Regions unlinked to the target region will not differ between the bulked samples of many individuals in BSA.

The markers of the present invention can be used to isolate or substantially purify an allele of a quantitative trait locus that is also located on linkage group U26

of a *Glycine soja* plant. Construction of an overlapping series of clones (a clone contig) across the region can provide the basis for a physical map encompassing an allele of a quantitative trait locus that are located on linkage group U26 of a *Glycine soja* plant. The yeast artificial chromosome (YAC) cloning system has facilitated chromosome walking and large-size cloning strategies. A sequence tag site (STS) content approach utilizing the markers of the present invention can be used for the construction of YAC clones across chromosome regions. Such an STS content approach to the construction of YAC maps can provide a detailed and ordered STS-based map of any chromosome region, including the region encompassing the allele of a quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant. YAC maps can be supplemented by detailed physical maps are constructed across the region by using BAC, PAC, or bacteriophage P1 clones that contain inserts ranging in size from 70 kb to several hundred kilobases (Cregan, *Theor. Appl.Gen. 78*:919-928 (1999); Sternberg, *Proc. Natl. Acad. Sci. 87:*103-107 (1990); Sternberg, *Trends Genet. 8:*11-16 (1992); Sternberg *et al., New Biol. 2:*151-162 (1990); Ioannou *et al., Nat. Genet. 6:*84-89 (1994); Shizuya *et al., Proc. Natl. Acad. Sci. 89:*8794-8797 (1992)).

Overlapping sets of clones are derived by using the available markers of the present invention to screen BAC, PAC, bacteriophage P1, or cosmid libraries. In addition, hybridization approaches can be used to convert the YAC maps into BAC, PAC, bacteriophage P1, or cosmid contig maps. Entire YACs and products of inter-*Alu*-PCR as well as primer sequences from appropriate STSs can be used to screen BAC, PAC, bacteriophage P1, or cosmid libraries. The clones isolated for any region can be assembled into contigs using STS content information and fingerprinting approaches (Sulston *et al., Comput. Appl. Biosci. 4:*125-132 (1988)).

The invention also provides a substantially purified nucleic acid molecule encoding a quantitative trait allele, where the allele is also located on linkage group U26 of a *Glycine soja* plant and preferably where the allele is also located on linkage group U26 of a *Glycine soja* plant between and including nucleic acid marker U3944117 and within 50 cM of U3944117 or its complement on linkage group U26.

The degeneracy of the genetic code, which allows different nucleic acid sequences to code for the same protein or peptide, is known in the literature. As used herein a nucleic acid molecule is degenerate of another nucleic acid molecule when the nucleic acid molecules encode for the same amino acid sequences but comprise

different nucleotide sequences. An aspect of the present invention is that the nucleic acid molecules of the present invention include nucleic acid molecules that are degenerate of the nucleic acid molecule that encodes the protein(s) of the quantitative trait alleles.

Another aspect of the present invention is that the nucleic acid molecules of the present invention include nucleic acid molecules that are homologues of the nucleic acid molecule that encodes the one or more of the proteins associated with the quantitative trait locus.

Exogenous genetic material may be transferred into a plant by the use of a DNA vector or construct designed for such a purpose. A particularly preferred subgroup of exogenous material comprises a nucleic acid molecule of the present invention. Design of such a vector is generally within the skill of the art (*See*, Plant Molecular Biology: A Laboratory Manual, eds. Clark, Springer, New York (1997)). Examples of such plants, include, without limitation, alfalfa, *Arabidopsis*, barley, *Brassica*, broccoli, cabbage, citrus, cotton, garlic, oat, oilseed rape, onion, canola, flax, maize, an ornamental plant, pea, peanut, pepper, potato, rice, rye, sorghum, soybean, strawberry, sugarcane, sugarbeet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, oil palm, *Phaseolus* etc.

A construct or vector may include a plant promoter to express the protein or protein fragment of choice. A number of promoters which are active in plant cells have been described in the literature. These include the nopaline synthase (NOS) promoter (Ebert *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 84*:5745-5749 (1987)), the octopine synthase (OCS) promoter (which are carried on tumor-inducing plasmids of *Agrobacterium tumefaciens*), the caulimovirus promoters such as the cauliflower mosaic virus (CaMV) 19S promoter (Lawton *et al.*, *Plant Mol. Biol. 9*:315-324 (1987)) and the CaMV 35S promoter (Odell *et al.*, *Nature 313*:810-812 (1985)), the figwort mosaic virus 35S-promoter, the light-inducible promoter from the small subunit of ribulose-1,5-bis-phosphate carboxylase (ssRUBISCO), the Adh promoter (Walker *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 84*:6624-6628 (1987)), the sucrose synthase promoter (Yang *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 87*:4144-4148 (1990)), the R gene complex promoter (Chandler *et al.*, *The Plant Cell 1*:1175-1183 (1989)), and the chlorophyll a/b binding protein gene promoter, etc. These promoters have

been used to create DNA constructs which have been expressed in plants; *see, e.g.*, PCT publication WO 84/02913.

Promoters which are known or are found to cause transcription of DNA in plant cells can be used in the present invention. Such promoters may be obtained from a variety of sources such as plants and plant viruses. In addition to promoters that are known to cause transcription of DNA in plant cells, other promoters may be identified for use in the current invention by screening a plant cDNA library for genes which are selectively or preferably expressed in the target tissues or cells.

For the purpose of expression in source tissues of the plant, such as the leaf, seed, root or stem, it is preferred that the promoters utilized in the present invention have relatively high expression in these specific tissues. For this purpose, one may choose from a number of promoters for genes with tissue- or cell-specific or -enhanced expression. Examples of such promoters reported in the literature include the chloroplast glutamine synthetase GS2 promoter from pea (Edwards *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 87*:3459-3463 (1990)), the chloroplast fructose-1,6-biphosphatase (FBPase) promoter from wheat (Lloyd *et al.*, *Mol. Gen. Genet. 225*:209-216 (1991)), the nuclear photosynthetic ST-LS1 promoter from potato (Stockhaus *et al.*, *EMBO J. 8*:2445-2451 (1989)), the serine/threonine kinase (PAL) promoter and the glucoamylase (CHS) promoter from *Arabidopsis thaliana*. Also reported to be active in photosynthetically active tissues are the ribulose-1,5-bisphosphate carboxylase (RbcS) promoter from eastern larch (*Larix laricina*), the promoter for the *Cab* gene, *cab6*, from pine (Yamamoto *et al.*, *Plant Cell Physiol. 35*:773-778 (1994)), the promoter for the *Cab-1* gene from wheat (Fejes *et al.*, *Plant Mol. Biol. 15*:921-932 (1990)), the promoter for the *Cab-1* gene from spinach (Lubberstedt *et al.*, *Plant Physiol. 104*:997-1006 (1994)), the promoter for the *Cab1R* gene from rice (Luan *et al.*, *Plant Cell. 4*:971-981 (1992)), the pyruvate, orthophosphate dikinase (PPDK) promoter from *Zea mays* (Matsuoka *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 90*:9586-9590 (1993)), the promoter for the tobacco Lhcb1*2 gene (Cerdan *et al.*, *Plant Mol. Biol. 33*:245-255. (1997)), the *Arabidopsis thaliana* SUC2 sucrose-H+ symporter promoter (Truernit *et al.*, *Planta. 196*:564-570 (1995)), and the promoter for the thylakoid membrane genes from spinach (psaD, psaF, psaE, PC, FNR, atpC, atpD, cab, rbcS). Other promoters for the chlorophyll a/b-binding genes may also be utilized in the present invention, such as the promoters for LhcB

gene and PsbP gene from white mustard (*Sinapis alba*; Kretsch *et al.*, *Plant Mol. Biol. 28*:219-229 (1995)).

For the purpose of expression in sink tissues of the plant, such as the tuber of the potato plant, the fruit of tomato, or the seed of *Zea mays*, wheat, rice, and barley, it is preferred that the promoters utilized in the present invention have relatively high expression in these specific tissues. A number of promoters for genes with tuber-specific or -enhanced expression are known, including the class I patatin promoter (Bevan *et al.*, *EMBO J. 8*:1899-1906 (1986); Jefferson *et al.*, *Plant Mol. Biol. 14*:995-1006 (1990)), the promoter for the potato tuber ADPGPP genes, both the large and small subunits, the sucrose synthase promoter (Salanoubat and Belliard, *Gene. 60*:47-56 (1987), Salanoubat and Belliard, *Gene. 84*:181-185 (1989)), the promoter for the major tuber proteins including the 22 kd protein complexes and proteinase inhibitors (Hannapel, *Plant Physiol. 101*:703-704 (1993)), the promoter for the granule bound starch synthase gene (GBSS) (Visser *et al.*, *Plant Mol. Biol. 17*:691-699 (1991)), and other class I and II patatins promoters (Koster-Topfer *et al.*, *Mol Gen Genet. 219*:390-396 (1989); Mignery *et al.*, *Gene. 62*:27-44 (1988)).

Other promoters can also be used to express a protein in specific tissues, such as seeds or fruits. The promoter for β-conglycinin (Chen *et al.*, *Dev. Genet. 10*: 112-122 (1989)) or other seed-specific promoters such as the napin and phaseolin promoters, can be used. The zeins are a group of storage proteins found in *Zea mays* endosperm. Genomic clones for zein genes have been isolated (Pedersen *et al.*, *Cell 29*:1015-1026 (1982)), and the promoters from these clones, including the 15 kD, 16 kD, 19 kD, 22 kD, 27 kD, and gamma genes, could also be used. Other promoters known to function, for example, in *Zea mays* include the promoters for the following genes: *waxy, Brittle, Shrunken 2*, branching enzymes I and II, starch synthases, debranching enzymes, oleosins, glutelins, and sucrose synthases. A particularly preferred promoter for *Zea mays* endosperm expression is the promoter for the glutelin gene from rice, more particularly the Osgt-1 promoter (Zheng *et al.*, *Mol. Cell Biol. 13*:5829-5842 (1993)). Examples of promoters suitable for expression in wheat include those promoters for the ADPglucose phosphorylase (ADPGPP) subunits, the granule bound and other starch synthase, the branching and debranching enzymes, the embryogenesis-abundant proteins, the gliadins, and the glutenins. Examples of such promoters in rice include those promoters for the ADPGPP subunits, the granule bound and other starch synthase, the branching enzymes, the debranching enzymes,

sucrose synthases, and the glutelins. A particularly preferred promoter is the promoter for rice glutelin, Osgt-1. Examples of such promoters for barley include those for the ADPGPP subunits, the granule bound and other starch synthase, the branching enzymes, the debranching enzymes, sucrose synthases, the hordeins, the embryo globulins, and the aleurone specific proteins.

Root specific promoters may also be used. An example of such a promoter is the promoter for the acid chitinase gene (Samac *et al.*, *Plant Mol. Biol. 25*:587-596 (1994)). Expression in root tissue could also be accomplished by utilizing the root specific subdomains of the CaMV35S promoter that have been identified (Lam *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 86*:7890-7894 (1989)). Other root cell specific promoters include those reported by Conkling *et al.*, *Plant Physiol. 93*:1203-1211 (1990).

Additional promoters that may be utilized are described, for example, in U.S. Patent Nos. 5,378,619, 5,391,725, 5,428,147, 5,447,858, 5,608,144, 5,608,144, 5,614,399, 5,633,441, 5,633,435, and 4,633,436. In addition, a tissue specific enhancer may be used (Fromm *et al.*, *The Plant Cell 1*:977-984 (1989)).

Constructs or vectors may also include with the coding region of interest a nucleic acid sequence that acts, in whole or in part, to terminate transcription of that region. For example, such sequences have been isolated including the Tr7 3' sequence and the NOS 3' sequence (Ingelbrecht *et al.*, *The Plant Cell 1*:671-680 (1989); Bevan *et al.*, *Nucleic Acids Res. 11*:369-385 (1983)), or the like.

A vector or construct may also include regulatory elements. Examples of such include the Adh intron 1 (Callis *et al.*, *Genes and Develop. 1*:1183-1200 (1987)), the sucrose synthase intron (Vasil *et al.*, *Plant Physiol. 91*:1575-1579 (1989)) and the TMV omega element (Gallie *et al.*, *The Plant Cell 1*:301-311 (1989)). These and other regulatory elements may be included when appropriate.

A vector or construct may also include a selectable marker. Selectable markers may also be used to select for plants or plant cells that contain the exogenous genetic material. Examples of such include, but are not limited to, a neo gene (Potrykus *et al.*, *Mol. Gen. Genet. 199*:183-188 (1985)) which codes for kanamycin resistance and can be selected for using kanamycin, G418, etc.; a bar gene which codes for bialaphos resistance; a mutant EPSP synthase gene (Hinchee *et al.*, *Bio/Technology 6*:915-922 (1988)) which encodes glyphosate resistance; a nitrilase gene which confers resistance to bromoxynil (Stalker *et al.*, *J. Biol. Chem. 263*:6310-

6314 (1988)); a mutant acetolactate synthase gene (ALS) which confers imidazolinone or sulphonylurea resistance (European Patent Application 154,204 (Sept. 11, 1985)); and a methotrexate resistant DHFR gene (Thillet *et al., J. Biol. Chem. 263*:12500-12508 (1988)).

A vector or construct may also include a transit peptide. Incorporation of a suitable chloroplast transit peptide may also be employed (European Patent Application Publication Number 0218571). Translational enhancers may also be incorporated as part of the vector DNA. DNA constructs could contain one or more 5' non-translated leader sequences which may serve to enhance expression of the gene products from the resulting mRNA transcripts. Such sequences may be derived from the promoter selected to express the gene or can be specifically modified to increase translation of the mRNA. Such regions may also be obtained from viral RNAs, from suitable eukaryotic genes, or from a synthetic gene sequence. For a review of optimizing expression of transgenes, see Koziel *et al., Plant Mol. Biol. 32*:393-405 (1996).

A vector or construct may also include a screenable marker. Screenable markers may be used to monitor expression. Exemplary screenable markers include a β-glucuronidase or uidA gene (GUS) which encodes an enzyme for which various chromogenic substrates are known (Jefferson, *Plant Mol. Biol, Rep. 5:*387-405 (1987); Jefferson *et al., EMBO J. 6:*3901-3907 (1987)); an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta *et al.*, Stadler Symposium *11*:263-282 (1988)); a β-lactamase gene (Sutcliffe *et al., Proc. Natl. Acad. Sci. (U.S.A.) 75*:3737-3741 (1978)), a gene which encodes an enzyme for which various chromogenic substrates are known (*e.g.*, PADAC, a chromogenic cephalosporin); a luciferase gene (Ow *et al., Science 234*:856-859 (1986)); a xylE gene (Zukowsky *et al., Proc. Natl. Acad. Sci. (U.S.A.) 80*:1101-1105 (1983)) which encodes a catechol dioxygenase that can convert chromogenic catechols; an α-amylase gene (Ikatu *et al., Bio/Technol. 8*:241-242 (1990)); a tyrosinase gene (Katz *et al., J. Gen. Microbiol. 129*:2703-2714 (1983)) which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to melanin; and an α-galactosidase.

Included within the terms "selectable or screenable marker genes" are also genes which encode a secretable marker whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers which

encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes which can be detected catalytically. Secretable proteins fall into a number of classes, including small, diffusible proteins which are detectable, (*e.g.*, by ELISA), small active enzymes which are detectable in extracellular solution (*e.g.*, α-amylase, β-lactamase, phosphinothricin transferase), or proteins which are inserted or trapped in the cell wall (such as proteins which include a leader sequence such as that found in the expression unit of extension or tobacco PR-S). Other possible selectable and/or screenable marker genes will be apparent to those of skill in the art.

There are many methods for introducing transforming nucleic acid molecules into plant cells. Suitable methods are believed to include virtually any method by which nucleic acid molecules may be introduced into a cell, such as by *Agrobacterium* infection or direct delivery of nucleic acid molecules such as, for example, by PEG-mediated transformation, by electroporation or by acceleration of DNA coated particles, etc. (Potrykus, *Ann. Rev. Plant Physiol. Plant Mol. Biol. 42*:205-225 (1991); Vasil, *Plant Mol. Biol. 25*:925-937 (1994)). For example, electroporation has been used to transform *Zea mays* protoplasts (Fromm *et al.*, *Nature 312*:791-793 (1986)).

Other vector systems suitable for introducing transforming DNA into a host plant cell include but are not limited to binary artificial chromosome (BIBAC) vectors (Hamilton *et al.*, *Gene 200*:107-116 (1997)), and transfection with RNA viral vectors (Della-Cioppa *et al.*, *Ann. N.Y. Acad. Sci.* (1996), 792 (Engineering Plants for Commercial Products and Applications), 57-61).

Technology for introduction of DNA into cells is well known to those of skill in the art. Four general methods for delivering a gene into cells have been described: (1) chemical methods (Graham and van der Eb, *Virology* 54:536-539 (1973)); (2) physical methods such as microinjection (Capecchi, *Cell* 22:479-488 (1980)), electroporation (Wong and Neumann, *Biochem. Biophys. Res. Commun. 107*:584-587 (1982); Fromm *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 82*:5824-5828 (1985); U.S. Patent No. 5,384,253); and the gene gun (Johnston and Tang, *Methods Cell Biol. 43*:353-365 (1994)); (3) viral vectors (Clapp, *Clin. Perinatol. 20*:155-168 (1993); Lu *et al.*, *J. Exp. Med. 178*:2089-2096 (1993); Eglitis and Anderson, *Biotechniques 6*:608-614 (1988)); and (4) receptor-mediated mechanisms (Curiel *et al.*, *Hum. Gen. Ther. 3*:147-154 (1992), Wagner *et al.*, *Proc. Natl. Acad. Sci. USA 89*:6099-6103 (1992)).

Acceleration methods that may be used include, for example, microprojectile bombardment and the like. One example of a method for delivering transforming nucleic acid molecules to plant cells is microprojectile bombardment. This method has been reviewed by Yang and Christou, eds., *Particle Bombardment Technology for Gene Transfer*, Oxford Press, Oxford, England (1994). Non-biological particles (microprojectiles) that may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold, platinum, and the like.

A particular advantage of microprojectile bombardment, in addition to it being an effective means of reproducibly transforming monocots, is that neither the isolation of protoplasts (Cristou *et al.*, *Plant Physiol. 87*:671-674 (1988)) nor the susceptibility of *Agrobacterium* infection are required. An illustrative embodiment of a method for delivering DNA into *Zea mays* cells by acceleration is a biolistics α-particle delivery system, which can be used to propel particles coated with DNA through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with corn cells cultured in suspension. Gordon-Kamm *et al.*, describes the basic procedure for coating tungsten particles with DNA (Gordon-Kamm *et al.*, *Plant Cell 2*:603-618 (1990)). The screen disperses the tungsten nucleic acid particles so that they are not delivered to the recipient cells in large aggregates. A particle delivery system suitable for use with the present invention is the helium acceleration PDS-1000/He gun is available from Bio-Rad Laboratories (Bio-Rad, Hercules, California)(Sanford *et al.*, *Technique 3*:3-16 (1991)).

For the bombardment, cells in suspension may be concentrated on filters. Filters containing the cells to be bombarded are positioned at an appropriate distance below the microprojectile stopping plate. If desired, one or more screens are also positioned between the gun and the cells to be bombarded.

Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the microprojectile stopping plate. If desired, one or more screens are also positioned between the acceleration device and the cells to be bombarded. Through the use of techniques set forth herein one may obtain up to 1000 or more foci of cells transiently expressing a marker gene. The number of cells in a focus which express the exogenous gene product 48 hours post-bombardment often range from one to ten and average one to three.

In bombardment transformation, one may optimize the pre-bombardment culturing conditions and the bombardment parameters to yield the maximum numbers of stable transformants. Both the physical and biological parameters for bombardment are important in this technology. Physical factors are those that involve manipulating the DNA/microprojectile precipitate or those that affect the flight and velocity of either the macro- or microprojectiles. Biological factors include all steps involved in manipulation of cells before and immediately after bombardment, the osmotic adjustment of target cells to help alleviate the trauma associated with bombardment, and also the nature of the transforming DNA, such as linearized DNA or intact supercoiled plasmids. It is believed that pre-bombardment manipulations are especially important for successful transformation of immature embryos.

In another alternative embodiment, plastids can be stably transformed. Method disclosed for plastid transformation in higher plants include on particle gun delivery of DNA containing a selectable marker and targeting of the DNA to the plastid genome through homologous recombination (Svab *et al. Proc. Natl. Acad. Sci. (U.S.A.) 87*:8526-8530 (1990): Svab and Maliga *Proc. Natl. Acad. Sci. (U.S.A.) 90*:913-917 (1993)); (Staub, J. M. and Maliga, P. *EMBO J. 12*:601-606 (1993), U.S. Patents 5, 451,513 and 5,545,818).

Accordingly, it is contemplated that one may wish to adjust various aspects of the bombardment parameters in small scale studies to fully optimize the conditions. One may particularly wish to adjust physical parameters such as gap distance, flight distance, tissue distance, and helium pressure. One may also minimize the trauma reduction factors by modifying conditions which influence the physiological state of the recipient cells and which may therefore influence transformation and integration efficiencies. For example, the osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells may be adjusted for optimum transformation. The execution of other routine adjustments will be known to those of skill in the art in light of the present disclosure.

Agrobacterium-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. See, for example the methods described by Fraley *et al., Bio/Technology 3*:629-635 (1985) and Rogers *et al., Methods Enzymol. 153*:253-

277 (1987). Further, the integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences, and intervening DNA is usually inserted into the plant genome as described (Spielmann *et al., Mol. Gen. Genet. 205*:34 (1986)).

Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing for convenient manipulations as described (Klee *et al.*, In: *Plant DNA Infectious Agents*, Hohn and Schell, eds., Springer-Verlag, New York, pp. 179-203 (1985). Moreover, technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes (Rogers *et al., Methods Enzymol. 153*:253-277 (1987)). In addition, *Agrobacterium* containing both armed and disarmed Ti genes can be used for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

A transgenic plant formed using *Agrobacterium* transformation methods typically contains a single gene on one chromosome. Such transgenic plants can be referred to as being hemizygous for the added gene. More preferred is a transgenic plant that is homozygous for the added structural gene; *i.e.*, a transgenic plant that contains two added genes, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) an independent segregant transgenic plant that contains a single added gene, germinating some of the seed produced and analyzing the resulting plants produced for the gene of interest.

It is also to be understood that two different transgenic plants can also be mated to produce offspring that contain two independently segregating added, exogenous genes. Selfing of appropriate progeny can produce plants that are homozygous for both added, exogenous genes that encode a polypeptide of interest. Back-crossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated, as is vegetative propagation.

Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and

combinations of these treatments (See for example, Potrykus *et al., Mol. Gen. Genet. 205*:193-200 (1986); Lorz *et al., Mol. Gen. Genet. 199*:178 (1985); Fromm *et al., Nature 319*:791 (1986); Uchimiya *et al., Mol. Gen. Genet. 204*:204 (1986); Marcotte *et al., Nature 335*:454-457 (1988))

Application of these systems to different plant strains depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration of cereals from protoplasts are described (Fujimura *et al., Plant Tissue Culture Letters 2*:74 (1985); Toriyama *et al., Theor Appl. Genet. 205*:34 (1986); Yamada *et al., Plant Cell Rep. 4*:85 (1986); Abdullah *et al., Biotechnology 4*:1087 (1986)).

To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil, *Bio/Technology 6*:397 (1988)). In addition, "particle gun" or high-velocity microprojectile technology can be utilized (Vasil *et al., Bio/Technology 10*:667 (1992)).

Using the latter technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small metal particles as described (Klein *et al., Nature 328:*70 (1987); Klein *et al., Proc. Natl. Acad. Sci. (U.S.A.) 85*:8502-8505 (1988); McCabe *et al., Bio/Technology 6:923* (1988)). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants.

Other methods of cell transformation can also be used and include but are not limited to introduction of DNA into plants by direct DNA transfer into pollen (Hess *et al., Intern Rev. Cytol. 107:*367 (1987); Luo *et al., Plant Mol Biol. Reporter 6:*165 (1988)), by direct injection of DNA into reproductive organs of a plant (Pena *et al., Nature 325*:274 (1987)), or by direct injection of DNA into the cells of immature embryos followed by the rehydration of desiccated embryos (Neuhaus *et al., Theor. Appl. Genet. 75*:30 (1987)).

The regeneration, development, and cultivation of plants from single plant protoplast transformants or from various transformed explants is well known in the art (Weissbach and Weissbach, In: *Methods for Plant Molecular Biology,* (Eds.), Academic Press, Inc. San Diego, CA, (1988)). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the

rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil.

The development or regeneration of plants containing the foreign, exogenous gene that encodes a protein of interest is well known in the art. Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important lines. Conversely, pollen from plants of these important lines is used to pollinate regenerated plants. A transgenic plant of the present invention containing a desired polypeptide is cultivated using methods well known to one skilled in the art.

There are a variety of methods for the regeneration of plants from plant tissue. The particular method of regeneration will depend on the starting plant tissue and the particular plant species to be regenerated.

Methods for transforming dicots, primarily by use of *Agrobacterium tumefaciens*, and obtaining transgenic plants have been published for cotton (U.S. Patent No. 5,004,863, U.S. Patent No. 5,159,135, U.S. Patent No. 5,518,908); soybean (U.S. Patent No. 5,569,834, U.S. Patent No. 5,416,011, McCabe *et. al., Bio/Technology 6:*923 (1988), Christou *et al., Plant Physiol. 87:*671-674 (1988)); *Brassica* (U.S. Patent No. 5,463,174); peanut (Cheng *et al., Plant Cell Rep. 15:*653-657 (1996), McKently *et al., Plant Cell Rep. 14:*699-703 (1995)); papaya; and pea (Grant *et al., Plant Cell Rep. 15:*254-258, (1995)).

Transformation of monocotyledons using electroporation, particle bombardment, and *Agrobacterium* have also been reported. Transformation and plant regeneration have been achieved in asparagus (Bytebier *et al., Proc. Natl. Acad. Sci. (USA) 84:*5354, (1987)); barley (Wan and Lemaux, *Plant Physiol 104:*37 (1994)); *Zea mays* (Rhodes *et al., Science 240:*204 (1988), Gordon-Kamm *et al., Plant Cell* 2:603-618 (1990), Fromm *et al., Bio/Technology 8:*833 (1990), Koziel *et al., Bio/Technology 11:*194, (1993), Armstrong *et al., Crop Science 35:*550-557 (1995)); oat (Somers *et al., Bio/Technology 10:*1589 (1992)); orchard grass (Horn *et al., Plant Cell Rep. 7:*469 (1988)); rice (Toriyama *et al., Theor Appl. Genet. 205:*34, (1986); Part *et al., Plant Mol. Biol. 32:*1135-1148, (1996); Abedinia *et al., Aust. J. Plant Physiol. 24:*133-141 (1997); Zhang and Wu, *Theor. Appl. Genet. 76:*835 (1988); Zhang *et al. Plant Cell Rep. 7:*379, (1988); Battraw and Hall, *Plant Sci. 86:*191-202 (1992); Christou *et al.,*

Bio/Technology 9:957 (1991)); rye (De la Pena *et al., Nature 325*:274 (1987)); sugarcane (Bower and Birch, *Plant J. 2*:409 (1992)); tall fescue (Wang *et al., Bio/Technology 10*:691 (1992)), and wheat (Vasil *et al., Bio/Technology 10*:667 (1992); U.S. Patent No. 5,631,152).

Assays for gene expression based on the transient expression of cloned nucleic acid constructs have been developed by introducing the nucleic acid molecules into plant cells by polyethylene glycol treatment, electroporation, or particle bombardment (Marcotte *et al., Nature 335*:454-457 (1988); Marcotte *et al., Plant Cell 1*:523-532 (1989); McCarty *et al., Cell 66*:895-905 (1991); Hattori *et al., Genes Dev. 6*:609-618 (1992); Goff *et al., EMBO J. 9*:2517-2522 (1990)). Transient expression systems may be used to functionally dissect gene constructs (*see generally*, Mailga *et al., Methods in Plant Molecular Biology,* Cold Spring Harbor Press (1995)). It is understood that any of the nucleic acid molecules of the present invention can be introduced into a plant cell in a permanent or transient manner in combination with other genetic elements such as vectors, promoters, enhancers etc.

In addition to the above discussed procedures, practitioners are familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of macromolecules (*e.g.*, DNA molecules, plasmids, etc.), generation of recombinant organisms and the screening and isolating of clones, (see for example, Sambrook *et al., Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989); Mailga *et al., Methods in Plant Molecular Biology,* Cold Spring Harbor Press (1995); Birren *et al., Genome Analysis: Detecting Genes,* 1, Cold Spring Harbor, New York (1998); Birren *et al., Genome Analysis: Analyzing DNA,* 2, Cold Spring Harbor, New York (1998); *Plant Molecular Biology: A Laboratory Manual,* eds. Clark, Springer, New York (1997)).

Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the present invention, unless specified.

Example 1

Two leaf discs are collected (approximately 40 mg) from a healthy leaf of a young *Glycine max* or *Glycine soja* plant and stored on wet ice or at 4°C. Tissue samples are then freeze-dried and stored at -20°C or -80°C. The frozen samples are kept as dry as possible and sealed from contact with the atmosphere. The freeze-dried

samples from
-20°C or -80°C, are allowed to warm up to room temperature prior to unsealing or opening. One leaflet (or 2 leaf discs) is inserted into an 1.5 ml Eppendorf tube, placed on dry ice, and crushed with a wooden dowel. Approximately 200 µl of microprep buffer (25 ml extraction buffer (350 mM sorbitol, 100 mM Tris-base, 5 mM EDTA-Na_2), 25 ml nuclei lysis buffer (1M Tris/HCl, 0.5 M EDTA, 5 M NaCl, 2% CTAB), 10 ml 5% sarkosyl, 0.1g Na bisulfite) is added to each sample. The sample is then homogenized. An additional 550 µl of microprep buffer is added, mixed by vortex for about 30-60 seconds, and incubated at 65°C for about 60 minutes. About 700 µl chloroform/isoamyl alcohol (24:1) is added, mixed well for about 10-30 seconds. Centrifugation of the tubes is performed at approximately 10,000 rpm for 5 minutes in a microcentrifuge. The aqueous phase is transferred into a new tube and RNA is removed from the extract by the addition of 30 µl of RNase (10 mg/ml) to the aqueous phase and incubated for 1 hour at room temperature. Approximately 500 µl ice-cold isopropanol is added to the aqueous extract, and the tubes inverted until the DNA precipitated. The precipitated solution is kept at 4°C for about 1 hour or overnight. Centrifugation of the tubes is performed at approximately 10,000 rpm for 5 minutes in a microcentrifuge. The supernatant is discarded and the pellet washed 1-3 times with 200 µl 70% ethanol. The ethanol is removed using a micropipette and pellet dried at 37°C for 10 minutes. The DNA is dissolved in 50 µl TE (10 mM Tris-HCL pH8.0, 0.1 mM EDTA), then kept overnight at 4°C. Centrifugation of the tubes is performed at approximately 10,000 rpm for 5 minutes and then the supernatant is transferred into new tubes. Using this method approximately 2 µg of DNA per mg of fresh leaf tissue is extracted.

The amount of DNA recovered is quantified by performing agarose gel electrophoresis on aliquots of the DNA extracted from the samples. The agarose gel is prepared as follows: 4 g agarose melted 400 ml 1X TBE (89 mM Tris-HCl, 89 mM boric acid, 2 mM EDTA), cooled to ~70°C and then 10 µl of 10 mg/ml ethidium bromide is added to the gel. A gel mold with comb for sample application is prepared and molten agarose poured into the mold. After the gel has solidified it is transferred to the electrophoresis apparatus containing approximately 2 L of 1X TBE buffer. For each sample, 9 µl (1 µl sample, 1 µl loading buffer with marker dye (50% glycerol, 0.1M EDTA, 0.1% bromophenol blue), 7 µl TE) is loaded. Molecular weight standards are included in the gel. The electrophoresis is conducted at approximately

100 mA for 2-4 hrs. The DNA concentration in each sample is estimated by it's staining intensity relative to the standards.

For analysis of the AFLP marker U3944117, the volume of the sample extract is adjusted with 1 X TE such that the concentration of the DNA in each samples is about 100-125 ng/μl. Three sets of restriction endonuclease digestions are performed: reaction 1. *Eco*RI, reaction 2. *Mse*I, and reaction 3. *Eco*RI/*Mse*I. 100-500 ng DNA is used per enzyme restriction digestion (Reaction 1: 8 μl of 5X RL buffer (6 ml water, 2.5 ml 1M KOAc, 500 μl 1 M Tris-HCl pH 8.0, 500 μl 1M MgOAc, 250 μl 1M DTT, 250 μl 10 mg/ml BSA), 31.5 μl water and DNA, and 0.5 μl *Eco*R1 (10 units/μl), incubated at 37°C for 3 or more hours or until DNA is determined to be completely digested; reaction 2 digestion: 8 μl of 5X RL buffer, 30.75 μl water and DNA, and 1.25 μl *Mse*I (5 units/μl) incubate at 37°C for 3 or more hours or until DNA is determined to be completely digested; reaction 3 digestion: 8 μl of 5X RL buffer, 30.25 μl water and DNA, 0.5 μl *Eco*RI, and 1.25 μl *Mse*I, incubated at 37°C for 3 or more hours or until DNA is determined to be completely digested. A 10 μl volume of the sample plus 2 μl of loading dye is loaded onto the agarose gel to determine whether the DNA is completely or partially digested. For set 1 (*Eco*RI), electrophoresis of the sample in the agarose gel is carried out until the dye reaches approximately 0.5 inches from the end of the gel, for set 2 (*Mse*I), until the dye is approximately 1 inch from the end. The digested DNA is observed under ultraviolet light, if all the DNA samples are completely digested, then the ligation reaction is carried out using the products of reaction 3 (*Eco*RI/*Mse*I). If some of the DNA samples are partially digested, these DNA samples are precipitated with 95% ethanol (2 volumes) or isopropanol (3/4 volume) and centrifuged in a microcentrifuge for 5 minutes, resuspended DNA in 1X TE, and the concentration is determined by electrophoresis on an agarose gel as described.

The addition of an *Eco*RI and *Mse*I adapter (Genosys Biotechnologies, Inc., Texas) to the *Eco*RI/*Mse*I digested DNA is performed using T4 DNA ligase. For each digested DNA reaction 4.8 μl of water, 2 μl of 5X RL buffer, 1 μl (5 pmol) *Eco*RI adapter, 1 μl (50 pmol) *Mse*I adapter, 1 μl 10mM ATP and 0.2 μl T4 DNA ligase (5 U/μl) is added and incubated at 37°C for between 3 hours and overnight. Centrifugation of the reaction tubes for a few seconds in a microcentrifuge is performed several times during the incubation period to coalesce the condensation. After ligation, half of each sample is diluted approximately 1:10 in TE. The dilution

ratio is adjusted based on the observed concentration of DNA after electrophoresis in the agarose gel, such that the final DNA concentration is similar in all samples.

For each sample a 5 µl aliquot is placed into a PCR tube, to which is added 36.4 µl of water, 5 µl of 10X PCR buffer, 1.5 µl of E39 primer (SEQ ID NO:15; 50 pmol), 1.5 µl of M44 primer (SEQ ID NO:16; 50 pmol), 0.4 µl of dNTPs (25mM), and 0.2 µl Taq polymerase (5 U/µl). The thermal cycler reaction conditions are 95°C 9 min.; 94°C 30 sec., 56°C 1 min., 72°C 1 min., 20 cycles; 4°C hold. An aliquot of each sample is checked by agarose gel electrophoresis. An aliquot of each of the preamplification samples is diluted approximately 1:20 in TE. The dilution ratio is adjusted according to the estimated DNA concentration such that the concentration is similar among all samples.

The E39 primer (SEQ ID NO:15; 20-30 ng) is radioactively labeled using T4 polynucleotide kinase and γ^{33}PdATP. Enough radioactive primer for 100 samples can be prepared by using 24 µl water, 5 µl T4 PNK buffer (250 mM Tris/HCl, 100 mM $MgCl_2$, 50 mM DTT, 5 mM spermidine), 1 µl E39 primer (SEQ ID NO:15; 25 ng/µl), 1 µl T4 PNK (10 U/ml), and 10 µl γ^{33}dATP (2000 Ci/mmol, 50 pmol). The primer reaction is incubated at 37°C for 60 minutes, then 10 minutes at 70°C and stored at -20°C. 5 µL diluted pre-amplification DNA (1:20 in TE) is added to each PCR tube and centrifuged. To each tube is added 11.66 µl of water, 2 µl of 10X PCR buffer, 0.6 µl of M44 primer (SEQ ID NO:16), 0.2 µl of cold E39 primer (SEQ ID NO:15; 25 ng/µl) 0.16 µl of dNTPs (25mM), 0.08 µl Taq polymerase (Perkin Elmer 5 U/µl), 0 and 0.5 µl of radiolabeled E39 primer (SEQ ID NO:15). The thermocycler conditions are 95°C 9 min.; 94°C 30 sec., 65°C 30 sec., (lower 0.7°C each cycle), 72°C 1 min, 13 cycles; 94°C 30 sec., 56°C 30 sec., 72°C 1 min., 23 cycles; 4°C hold. After the PCR reaction, 20 µL of formamide dye (98% formamide, 10 mM EDTA, 1 mg/ml xylene cyanol, 1 mg/ml bromophenol blue) is added to each tube. An additional reaction to denature the sample is carried out at thermocycler conditions 90°C 3 min., 4°C hold.

*Eco*RI primers are labeled with γ ^{33}P dATP as described above. Load 5 µl diluted preamplification DNA (1:20 in TE) is added into each PCR tube. *Mse*I primers (Genosys Biotechnologies, Inc., 2) and dNTPs (25mM) in 7 µl are added to each tube. To each reaction 5.42 µl water, 2 µl 10X PCR buffer, 0.2 µl E39 primer, 0.08 µl Taq polymerase (5 U/µl) and 0.5 µl of ^{33}P radiolabeled E39 primer (SEQ ID NO:15; 0.87 pmol), 8 µl into is added to each PCR tube.

Twenty to thirty ng primer DNA is labeled with ^{33}P using polynucleotide kinase method as previously described. For each PCR sample 16.4 µl water with 30 ng DNA, 2 µl 10X PCR buffer (Perkin Elmer, Cat# H0077), 0.3 µl E39 primer (SEQ ID NO:15, 50 pmol), 0.2 µl M44 primer (SEQ ID NO:16, 50 pmol), 0.5 µl dNTP mix (25 mM), 0.1 µl Taq polymerase (5 U/µl), 0.5 µl M44 (0.87 pmol) primer ^{33}P labeled.

An acrylamide gel is prepared using 56.5 ml water, 3.5 ml 10X TAE buffer, 10.5 ml 40% acrylamide stock solution, 50 µl TEMED, 0.06 g ammonium persulfate. To each PCR sample 20 µl of formamide loading dye is added to each sample and the samples are denatured at 90°C for 3 minutes with a 4°C hold in a thermocycler. 1.5 µl of each sample is loaded onto the gel. Gels are run at constant wattage to give a constant heat development during electrophoresis at 40 to 50 Volt/cm of the gel length. Gels should be run at approximately 50°C during electrophoresis. The electrophoresis is stopped when the Bromophenol blue dye is at the bottom of the gel. After electrophoresis, the gel is fixed for 30 minutes in 10% acetic acid, then rinsed for 10 minutes in tap-water. The gel is dried onto the glass plate using a hair-drier or 80°C oven. The gel is exposed to phospho-imaging screens for 18-24 hours. The exposed screens can be scanned with a Fuji BAS2000 or other suitable instrument following the manufacturers instructions and image saved for analysis.

Example 2

For analysis using the SSR markers the DNA extraction protocol is the same as described in Example 1 except the volume of the DNA sample is adjusted with 1 X TE such that the concentration of the DNA in each sample is about 1 ng/µl.

For each sample a 5 µl aliquot is placed into each well of a Perkin-Elmer MicroAmp Optical 96 Well reaction plates, to which is added 1.5 µl H20, 1.0 µl 10X PCR buffer, 0.04 µl 25 mM dNTPs, 1.0 µl Dye (20mM MgCl2, 20% sucrose, 1 mM Cresol Red), 1.5 µl 1µM mix of forward and reverse primers for each SSR marker, and 0.064 µl of 0.32 units of Taq polymerase. The marker pairs are SEQ ID NO. 17 and SEQ ID NO. 18 for SATT168; SEQ ID NO. 19 and SEQ ID NO 20 for SATT416; SEQ ID NO 21 and SEQ ID NO 22 for SAT_083; SEQ ID NO 23 and SEQ ID NO 24 for SATT474; SEQ ID NO 25 and SEQ ID NO 26 for SATT122; SEQ ID NO 27 and SEQ ID NO 28 for SATT556; SEQ ID NO 29 and SEQ ID NO 30 for Sct_094; SEQ ID NO 31 and SEQ ID NO 32 for SATT272; SEQ ID NO 33 and SEQ ID NO 34 for SATT020; SEQ ID NO 35 and SEQ ID NO 36 for SATT066; SEQ ID NO 37 and SEQ ID NO 38 for SATT534; SEQ ID NO 39 and SEQ ID NO 40 for SATT560.

Polymerase chain reaction is performed with the following thermal cycler conditions, 94°C 4 min.; 94°C 25 sec., 47°C 25 sec., 72°C 25 sec., 32 cycles; 72°C 3 min for final extension and 4°C hold.

An acrylamide gel is prepared using 56.5 ml water, 3.5 ml 10X TAE buffer, 10.5 ml 40% acrylamide stock solution, 50 µl TEMED, 0.06 g ammonium persulfate. A total of 5 µl of the PCR product is loaded onto the acrylamide gels on 1X TAE buffer. Molecular weight ladders are also loaded onto the gel to facilitate identification of SSR bands. Gels are run at for 45 minutes at 300V. The electrophoresis is stopped when the cresol red dye is at the bottom of the gel. Gels are then stained with SYBR green by mixing 20 µl of 10,000X SYBR green and 200 ml 1X TAE. The mixture should be enough to stain 20 gels. Gels are stained for 15-20 minutes with vigorous shaking. The gel bands are then visualized under a UV transilluminator. The PCR reaction product is then scored for the presence or absence of the bands on the appropriate molecular weights of SSR markers spanning the QTL.

Example 3

G. soja PI 407305 is from the Shanghai area of China belonging to maturity group 5. Crosses are made between soybean line HS-1 (Hartz Seed, Stuttgart, Arkansas) and *G. soja* accession PI 407305 (United Stated Department of Agriculture Soybean Germplasm Collection, University of Illinois, Urbana -Champaign, USA). Pollen from the F_1 progeny of that cross are then crossed back to parent line HS-1 to generate about 40 BC_1F_1 progeny. Each BC_1F_1 progeny is then grown and crossed again to parent line HS-1 to generate between 250 and 300 BC_2F_1 progeny. The BC_2F_1 progeny are grown and leaf samples are taken from each plant for subsequent DNA extraction and molecular marker genotyping. The BC_2F_1 plants are grown to maturity and BC_2F_2 seeds collected. BC_2F_2 seeds from each BC_2F_1 plant are then bulked. The resulting seeds from each of 266 BC_2F_1-derived progeny are used for yield trials in three locations: Jerseyville, Illinois, Stuttgart, Arkansas, and Rolling Forks, Mississippi.

The plots are laid out in a random split block design with a single replication, where blocks represent early, mid and late maturity groups to facilitate harvest. There are two-row 16-ft. plots, with the adapted parent, HS-1, as a border row on each side. Seeding rate is eight seeds per foot. Cultural practices such as herbicide applications and fertilization are carried out following the recommendations for soybean. For example in Jerseyville plots, Lasso is applied as pre-emergence herbicide at the rate of

3 qt/Acre and Fusilade is applied as post-emergence at the rate of 16 oz/Acre. At harvest, only the test rows are harvested and seed yield is adjusted to 13% moisture content to get the dry yield for each line using the formula: Dry yield = Actual yield x (1-% moisture at harvest)/(1-0.13). Seed yield per plot is converted into yield in bushels per acre using the formula: Plot size/Acre = lb/Acre.

For example, yield measured in lbs. from a 16-ft x 5 ft plot is converted to bushels per acre by multiplying it with a factor of 9.075. For the 1997 yield trials, the same experimental procedure is used except that there are two replications in each of the following locations: Evansville, IN, Stonington, IL, Marion AR, Galena, MD, Stuttgart, AR and Jerseyville, IL. In addition, there are two separate experiments in the Jerseyville, IL location. Lines are grown in high nitrogen (200lbs/A) and low nitrogen (0lbs/A) treatments to assess the effect of increased nitrogen input. For the 1998 yield trials, the same experimental procedure is used as in 1997 except only two locations, Stonington, IL and Jerseyville, IL are tested.

DNA marker analysis is performed among the BC_2F_1 plants. Leaf tissue is collected and DNA extracted from each of the 266 BC_2F_1 plants. Each line is genotyped with 212 AFLP markers and three morphological markers (seed color, pod color and growth habit) spanning the whole genome. A genetic map is generated mainly with amplified fragment length polymorphism (AFLP) markers using Mapmaker. AFLP analysis as described by Vos *et al., Nucleic Acids Res. 23*:4407-4414 (1995). Linkage between AFLP markers is inferred whenever the LOD score is 2.0 or higher. The LOD score is the $\log_{10}$ of the odds ratio between the odds of the null hypothesis (that the markers are linked) against an alternative hypothesis (that they are not linked). To locate putative yield-enhancing QTLs, significant association between marker locus and yield dare determined at P = 0.05 using Q-gene (Qgene, Version 2.23, C. Nelson, Cornell University) and SAS software (SAS Institute; http://www.sas.com). In some instances where a QTL falls below the given threshold for significance, the consistency of its effect on yield across years and locations is taken into consideration.

Both analysis of variance, single point and interval mapping analysis identified an AFLP marker locus, U3944117 on linkage group U26. U3944117 is correlated with a significantly higher seed yield by weight than the average of the progenies derived from BC_1F_1 population heterozygous for the wild and the cultivated alleles compared to those derived from plants homozygous for the cultivated alleles

Table 1 (1996 and 1997 Yield data) and Table 2 (1998 Yield Data). Most locations show a significant seed yield by weight increase on average in the progenies carrying the wild alleles at either locus, when compared with the progenies homozygous for the adapted alleles (LOD peak score 3.99, as judged by interval mapping, Table 3.). In all cases, the average percent yield increase of the plants carrying the alleles derived from PI407305 is statistically significant (Analysis of Variance) higher than that of the plants homozygous for the adapted alleles (Table 1 and Table 2). To facilitate the use of this exotic locus in improving yield of commercial cultivars the following procedure can be used. Briefly, a cross can be made with any of the progenies derived from the HS-1 x PI 407305 and derivatives thereof of PI407305 carrying the exotic locus with any potential cultivar that one wishes to improve. Using molecular marker analysis described earlier, one can monitor the positive transfer of the exotic yield-enhancing locus by checking the presence of the molecular marker band corresponding to U3944117 and SSR markers. Then a series of backcrosses (up to BC_5) to the commercial cultivar (recurrent parent) can be made to recover most of the agronomic properties of the recurrent parent. Prior to each backcross step, the positive transfer of the exotic alleles has to be validated among backcross-derived progenies (BCnFn) (where n=generation) using molecular marker analysis as previously described. The number of backcrosses depends on the level of recurrent parent recovery which can also be facilitated by the use of markers evenly distributed throughout the genome.

Besides increased yield, other phenotypic expressions of the yield QTL from PI4070305 can be observed. Increase in *Glycine max* plant height is a phenotypic marker of the QTL as shown in Table 4. When the *Glycine max* geneotype is homozygous for the QTL there is a significant (CV = 5) increase in plant height. The mean values shown in Table 4 are the averages of the height of the main stem of five plants in two replications of field grown plants. Plant height is a component of yield for soybean.

The number of pods on the soybean plant can also be a measure of yield for soybean. In Table 5 the number of pods on the main stem were counted on five plants in two replications from plants homozygous, heterozygous or negative for the QTL from PI4070305. This test does not show statistical significance (CV=17), however, there is a tendency for the QTL containing *Glycine max* to have about 25% more pods

on the main stem and this result is indicative of the yield QTL of the present invention.

Example 4

The United Stated Department of Agriculture Soybean Germplasm Collection, University of Illinois, Urbana -Champaign, USA contains approximately 10,000 Glycine max and 2,000 Glycine soja Pis. Such germplasm may be screened for the presence of an allele of a quantitative trait locus of the present invention. For example, marker analysis of approximately 100 *Glycine soja* PI's obtained from the USDA collection of soybean is conducted using the microsatellite sequences (SSRs) of Satt168, Satt416, Sat_083, Satt474, Satt122, Satt556, Sct_094, Satt272, Satt020 Satt066, Satt534, Satt560 or their complements and the methods of analysis for the presence of these markers in DNA extracts from tissue of *Glycine soja* PIs is described in Example 2. Table 6 shows that a only a small number of these *Glycine soja* PIs show the presence of only one of the above microsatellite marker sequences (Table 6).

Approximately 250 *Glycine max* PIs from the USDA collection of soybean are analyzed for the presence of the SSR markers Satt168, Satt416, Sat_083, Satt474, Satt122, Satt556, Sct_094, Satt272, Satt020 Satt066, Satt534, Satt560 or their complements using the methods of molecular plant breeding. Three of these *Glycine max* PIs are identified to each contain a single SSR marker (Table 7). Two PIs from Japan contain Satt020 and one line of unknown origin contains Satt556. The SSR markers used to identify the yield QTL from *Glycine soja* are infrequently present in *Glycine max* PIs.

None of the SSR markers is detected in the analysis of approximate 30 *Glycine max* elite lines.

In cases where the *Glycine soja* or *Glycine max* plants screened shares one or only a limited number of the markers associated with high yield in other *Glycine soja* or *max* plants such as PI407305, the presence or absence of an quantity trait locus in such plants can be confirmed by creating a mapping population and determining whether the progeny of such plants exhibit one of the physical traits, such as height, associated with the quantitative trait locus. The likelihood that any *Glycine soja* or *Glycine max* screened has a quantitative trait locus associated with yield increases as the number and the genomic colinearity (*i.e.* the degree that the order of the markers matches the order set forth in Figure 1) of the markers present increases.

Glycine max C83-1, C83-2, and C83-3 are sibling plants from the progeny of BC_2F_4 plants that are selfed. The presence of a molecular marker band corresponding to U3944117 is confirmed in *Glycine max* C83-1, C83-2 and C83-3. Seeds from sibling *Glycine max* plants C83-1, C83-2 and C83-3 were deposited with the American Type Culture Collection (ATCC, 10801 University Blvd, Manassas, Virginia, U.S.A., 20110-2209) on August 12, 1998 and assigned ATCC Nos. 203138, 203139 and 203140 respectively.

Glycine max C83-75 (deposited May 7, 1999 and assigned ATCC No. PTA-30) is a line from the progeny of BC_2F_4 selfed plants. This progeny of the HS-1 X PI40735 contains the U3944117 marker from *Glycine soja* associated with enhanced yield in *Glycine max.* The C83-75 is related to the C83-1, C83-2, and C83-3 lines. The C83-1, C83, C83-3, and C83-75 lines are useful for breeding the yield QTL identified by U3944117 into *Glycine max* varieties from all maturity groups.

TABLE 1

Yield data of Glycine max (HS-1) containing U3944117 marker for the high yield QTL from Glycine soja PI407305

Location	Year	#of Progeny lines having U3944117/# Tested	Percent Yield[2] Increase in lines having U3944117	P-Value
Stuttgart, AR	1996	39/256	11.0	0.0002
Jerseyville, IL	1996	41/264	5.0	<0.0001
Rolling Forks, MS	1996	38/232	10.0	0.03007
Stuttgart, AR	1997	19/222	26.0	0.0145
Jerseyville, IL	1997	29/73	5.3	ns
Jerseyville, IL(N)	1997	10/71	9.0	ns
Stonington, IL	1997	9/49	26	0.0474
Evansville, IN	1997	18/188	12.0	0.0012
Galena, MD	1997	20/153	15.0	0.0124
Combined Years and Locations			12.0	<0.0001

(N) high nitrogen
ns is not significant

TABLE 2

1998 Mean Yield Across Genotypes of Isogenic Populations Derived from HS-1 x PI407305 BC2 Mapping Population.

Genotype	Mean (bu/Ac)[2]	N[3]	Duncan Multiple range[1]
Homozygous QTL	44.552	40	A
Heterozygous QTL	43.267	50	A
QTL negative	37.635	18	B

[1]SAS grouping of statistically significant populations.

[2]Yield is measured as dry seed weight in bushels per acre.

[3]N is the number of lines tested

TABLE 3

Interval analysis of Linkage Group U26 Containing the Yield QTL derived from *Glycine soja* PI4070305

Marker	cM	N lines	F-value	P-value	LOD
Satt560	0	244	3.69	0.0559	1.9
Satt534	8.1	246	7.83	0.0055	3.51
Satt066	12.6	240	15.34	0.0001	3.67
U3944117	15	248	15.46	0.0001	2.85
Satt020	15.8	237	18.25	<0.0001	2.49
Satt272	15.8	250	16.58	0.0001	2.49
Sct_094	16.8	223	18.12	<0.0001	2.15
Satt556	17.3	240	13.97	0.0002	2.14
Satt122	17.3	253	13.76	0.0003	2.14
Satt474	17.3	256	14.01	0.0002	2.14
Sat_083	17.3	242	6.77	0.0098	2.14
Satt416	21.8	243	8.86	0.0032	1.66
Satt168	24.7	247	7.52	0.0066	1.16

TABLE 4

Comparison of Soybean Plant Height (cm) at Maturity

QTL Genotype	Mean*
Homozygous rep 1	65.93^{A}
Homozygous rep 2	65.03^{A}
Heterozygous rep 1	64.97^{AB}
Heterozygous rep 2	64.80^{AB}
QTL negative	58.87^{B}

*values with the same letters are not statistically significant (Duncan's multiple range test)

TABLE 5

Number of pods at main stem at Maturity

QTL geneotype	Mean*
Homozygous rep 1	40.33 A
Homozygous rep 2	43.47 A
Heterozygous rep 1	40.00 A
Heterozygous rep 2	41.00 A
QTL negative	30.53 A

* values with the same letters are not statistically significant (Duncan's multiple range test)

TABLE 6

SSR Markers Associated with *Glycine soja* PI's

U26 SSR marker	PI#	Geographic location
Satt168	549035A	China, Liaoning
Satt416	483464B	China, Ningxia
Satt416	468398A	China, Shanxi
Satt122	549048	China, Beijing
Satt556	479749	China, Jilin
Satt556	549034	China Liaoning
Satt272	522204	Russian Fed, Primorya
Satt272	507788	Russian Fed
Satt020	522220B	Russian Fed
Satt066	522200B	Russian Fed
Satt534	549037	China, Liaoning
Satt534	549032	China, Liaoning
Satt534	549036	China, Liaoning

TABLE 7

SSR Markers Associated with *Glycine max* PI's

U26 SSR marker	PI#	Geographic location
Satt020	209331	Japan
Satt020	426762	Japan
Satt556	578340B	

We Claim

1. A *Glycine max* plant having an allele of a quantitative trait locus associated with enhanced yield in said *Glycine max* plant, wherein said allele of said quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

2. The *Glycine max* plant according to claim 1, wherein said *Glycine max* plant is homozygous at said quantitative trait locus.

3. The *Glycine max* plant according to claim 1, wherein said *Glycine max* plant is heterozygous at said quantitative trait locus.

4. The *Glycine max* plant according to claim 1, wherein said *Glycine soja* plant is PI407305 or progeny thereof.

5. The *Glycine max* according to claim 1, wherein said allele of said quantitative trait locus located on said linkage group U26 of said *Glycine soja* plant is genetically linked to a complement of a marker nucleic acid, wherein said marker nucleic acid molecule is selected from the group consisting of a marker nucleic acid molecule in a region between Satt168 and Satt560.

6. The *Glycine max* according to claim 1, wherein said *Glycine max* plant is an elite plant.

7. The *Glycine max* according to claim 1, wherein said *Glycine max* plant exhibits an enhanced yield.

8. An elite *Glycine max* plant having an allele of a quantitative trait locus associated with enhanced yield in the elite *Glycine max* plant, wherein said allele of the quantitative trait locus is also located on linkage group U26 of an exotic *Glycine* plant.

9. The *Glycine max* according to claim 8, wherein said allele of said quantitative trait locus located on said linkage group U26 of said exotic *Glycine* plant is genetically linked to a complement of a marker nucleic acid, wherein said marker nucleic acid molecule is selected from the group consisting of a marker nucleic acid molecule in a region between Satt168 and Satt560.

10. The *Glycine max* according to claim 8, wherein said *Glycine max* plant is an elite plant.

11. The *Glycine max* according to claim 8, wherein said *Glycine max* plant exhibits an enhanced yield.

12. A *Glycine max* plant comprising DNA where said DNA has the same sequence as DNA found in an allele of a quantitative trait locus derived from *Glycine soja* PI407305 or progeny thereof.

13. An elite *Glycine max* plant comprising an allele of a quantitative trait locus derived from an exotic *Glycine* plant, wherein said quantitative trait locus is also located on linkage group U26 of *Glycine soja* PI407305.

14. A *Glycine max* plant having a genome, wherein said genome comprises a genetic locus having an allele of a quantitative trait locus genetically linked to the complement of marker nucleic acid molecule U3944117 or its complement.

15. The *Glycine max* plant according to claim 14, wherein said genetic locus is located between about 0 and about 50 centimorgans from said complement of said marker nucleic acid.

16. The *Glycine max* plant according to claim 15, wherein said genetic locus is located between about 0 and about 40 centimorgans from said complement of said marker nucleic acid.

17. The *Glycine max* plant according to claim 16, wherein said genetic locus is located between about 0 and about 25 centimorgans from said complement of said marker nucleic acid.

18. The *Glycine max* plant according to claim 17, wherein said genetic locus is located between about 0 and about 10 centimorgans from said complement of said marker nucleic acid.

19. The *Glycine max* plant according to claim 18, wherein said genetic locus is located between about 0 and about 5 centimorgans from said complement of said marker

20. The *Glycine max* plant according to claim 19, wherein said genetic locus is located between about 0 and about 3 centimorgans from said complement of said marker nucleic acid.

21. The *Glycine max* plant according to claim 14, wherein said marker nucleic acid molecule exhibits a LOD score for enhanced yield of greater than 2.0 for said allele.

22. The *Glycine max* plant according to claim 21, wherein said marker nucleic acid molecule exhibits a LOD score for enhanced yield of greater than 3.0 for said allele.

23. The *Glycine max* plant according to claim 22, wherein said marker nucleic acid molecule exhibits a LOD score for enhanced yield of greater than 3.5 for said allele.

24. The *Glycine max* plant according to claim 23, wherein said marker nucleic acid molecule exhibits a LOD score for enhanced yield of greater than 4.0 for said allele.

25. A *Glycine max* plant comprising an allele of a quantitative trait locus derived from *Glycine soja* PI407305 or progeny thereof, wherein said quantitative trait locus derived is from *Glycine soja* PI407305 or progeny thereof is located on linkage group U26.

26. The *Glycine max* plant according to claim 25, which is homozygous for a quantitative trait locus derived from *Glycine soja* PI407305 or progeny thereof.

27. The *Glycine max* plant according to claim 25, which is heterozygous for a quantitative trait locus derived from *Glycine soja* PI407305 or progeny thereof.

28. The *Glycine max* plant according to claim 25, wherein said progeny of said *Glycine soja* PI407305 has a *Glycine soja* nuclear genetic contribution of less than about 25%.

29. A *Glycine max* plant having a genome, wherein said genome has a least two polymorphisms capable of being detected by a polymorphic marker selected from the group consisting of: Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement, Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement.

30. The *Glycine max* plant having a genome according to claim 29, wherein said genome has at least four of said polymorphisms.

31. A container of over 40,000 *Glycine max* seeds, wherein over 80% of said seeds have an allele of a quantitative trait locus associated with enhanced yield in said *Glycine max* plant, wherein said allele of a quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

32. The container of over 40,000 *Glycine max* seeds according to claim 31, wherein said allele of a quantitative trait locus is derived from *Glycine soja* PI 407305.

33. A *Glycine max* plant, which exhibits an enhanced yield compared to a first parent, said *Glycine max* plant comprising a genome homozygous or heterozygous with respect to genetic alleles that are native to a second parent selected from the group consisting of *Glycine soja* PI407305 and progeny thereof and non-native to a first parent, wherein said first parent is an elite *Glycine max* plant.

34. An elite *Glycine max* plant, which exhibits an enhanced yield compared to a first parent, the elite *Glycine max* plant comprising a genome homozygous or heterozygous with respect to a genetic allele that is native to a second parent selected from the group consisting of an exotic *Glycine* plant having an allele of a quantitative trait locus, where the quantitative trait locus is also located on linkage group U26 of *Glycine soja* PI407305.

35. A *Glycine max* plant selected for by screening for an enhanced yield in said *Glycine max* plant, said selection comprising interrogating genomic DNA for the presence or absence of a marker molecule that is genetically linked to an allele of a quantitative trait locus associated with enhanced yield in said *Glycine max* plant, wherein said allele of a quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

36. A *Glycine max* seed selected from a *Glycine max* plant by screening for an enhanced yield in said *Glycine max* plant, said selection comprising interrogating genomic DNA for the presence of a marker molecule that is genetically linked to an allele of a quantitative trait locus associated with enhanced yield in said *Glycine max* plant, wherein said allele of said quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

37. An elite *Glycine max* plant selected for by screening for an enhanced yield in the *Glycine max* plant, the selection comprising interrogating genomic DNA for the presence of a marker molecule that is genetically linked to an allele of a quantitative trait locus associated with enhanced yield in an exotic *Glycine* plant, wherein the allele of a quantitative trait locus is also located on linkage group U26 of a *Glycine soja* plant.

38. A substantially purified marker nucleic acid molecule, said nucleic acid molecule capable of specifically hybridizing to a second nucleic acid molecule that is U3944117 or its complement.

39. A substantially purified nucleic acid molecule encoding a quantitative trait allele, wherein said allele is also located on linkage group U26 of a *Glycine* plant.

40. The substantially purified nucleic acid molecule encoding a quantitative trait allele according to claim 39, wherein said allele is also located on linkage group U26 of a *Glycine soja* plant between Satt168 and Satt560.

41. The substantially purified nucleic acid molecule encoding a quantitative trait locus according to claim 40, wherein said *Glycine soja* plant is *Glycine soja* PI407305.

42. A plant transformed with a DNA construct comprising a nucleic acid molecule according to claim 39, wherein said plant is selected from the group comprising: alfalfa, *Arabidopsis thaliana*, barley, *Brassica*, broccoli, cabbage, citrus, cotton, garlic, oat, oilseed rape, onion, canola, flax, an ornamental plant, pea, peanut, pepper, potato, rice, rye, soybean, sorghum, soybean, strawberry, sugarcane, sugarbeet, tomato, wheat, maize, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, oil palm, and *Phaseolus*.

43. A method for the production of an elite *Glycine max* plant having enhanced yield comprising:

(A) crossing a *Glycine soja* PI407305 plant or progeny thereof with a *Glycine max* plant to produce a segregating population;

(B) screening the segregating population for a member having an allele derived from *Glycine soja* PI407305 plant or progeny thereof that mapped to linkage group U26 of said *Glycine soja* PI407305 plant or progeny thereof, wherein said allele is associated with said enhanced yield in said *Glycine max* plant; and

(C) selecting the member for further crossing and selection, wherein said member selected has said allele derived from *Glycine soja* PI407305 plant or progeny thereof that mapped to linkage group U26

44. The method for the production of an elite *Glycine max* plant having enhanced yield according to claim 43, wherein said progeny of said *Glycine soja* PI407305 has a *Glycine soja* nuclear genetic contribution of less than about 25%.

45. A method of introgressing enhanced yield into a *Glycine max* plant comprising using a nucleic acid marker for marker assisted selection of said *Glycine max* plant, said nucleic acid marker complementary to a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant between Satt168 and Satt560, wherein the source of said enhanced yield is *Glycine soja* PI407305 or progeny thereof, and introgressing said enhanced yield into a *Glycine max* plant.

46. The method of introgressing enhanced yield into a *Glycine max* plant according to claim 45, wherein said introgression of said enhanced yield is carried out by backcrossing with a *Glycine max* recurrent parent.

47. A method for screening for enhanced yield comprising interrogating genomic DNA for the presence or absence of a marker molecule that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant between Satt168 and Satt560, wherein the source of said enhanced yield is *Glycine soja* PI407305 or progeny thereof; and detecting said presence or absence of said marker.

48. The method for screening for enhanced yield according to claim 47, wherein said marker molecule is a microsatellite marker.

49. The method for screening for enhanced yield according to claim 47, wherein said marker molecule is U3944117.

50. The method for screening for enhanced yield according to claim 47, wherein the marker molecule is detected by DNA amplification using a forward and a reverse primer.

51. The method for screening for enhanced yield according to claim 47, wherein said detecting of said presence or absence of said marker is detected by a detection method selected from the group consisting of AFLP, RFLP, RAPD, SNP and microsatellite analysis.

52. The method for screening for enhanced yield according to claim 47, wherein said marker exhibits a LOD for enhanced yield of greater than 2.0.

53. The method for screening for enhanced yield according to claim 52, wherein said marker exhibits a LOD for enhanced yield of greater than 3.5.

54. A method for determining the likelihood of a quantitative trait allele for enhanced yield in a *Glycine max* plant comprising the steps of:

(A) obtaining genomic DNA from said plant;

(B) detecting a marker molecule, wherein said marker molecule specifically hybridizes with a nucleic acid sequence that is genetically linked to a nucleic acid sequence that is located on linkage group U26 of a *Glycine soja* plant between Satt168 and Satt560; and

(C) determining the presence or absence of said marker molecule, wherein the presence or absence of said marker molecule is indicative of a quantitative trait allele for enhanced yield.

55. The method for determining the likelihood of a quantitative trait allele for enhanced yield in a *Glycine max* plant according to claim 54, wherein said marker molecule is a microsatellite marker.

56. The method for determining the likelihood of a quantitative trait allele for enhanced yield in a *Glycine max* plant according to claim 54, wherein said marker molecule is U3944117.

57. The method for determining the likelihood of a quantitative trait allele for enhanced yield in a *Glycine max* plant according to claim 54, wherein the marker molecule is detected by DNA amplification using a forward and a reverse primer.

58. The method for determining the likelihood of a quantitative trait allele for enhanced yield in a *Glycine max* plant according to claim 54, wherein said detecting of said presence or absence of said marker is detected by a detection method selected from the group consisting of AFLP, RFLP, RAPD and microsatellite analysis.

59. The method for determining the likelihood of a quantitative trait allele for enhanced yield in a *Glycine max* plant according to claim 54, wherein said marker exhibits a LOD for enhanced yield of greater than 2.0.

60. The method for determining the likelihood of a quantitative trait allele for enhanced yield in a *Glycine max* plant according to claim 55, wherein said marker exhibits a LOD for enhanced yield of greater than 3.5.

61. A method for determining the probability that a plant has a quantitative trait allele for enhanced yield:

(A) detecting the level, presence or absence of a polymorphism genetically linked to a quantitative trait allele for enhanced yield, wherein said polymorphism is located on linkage group U26 of a *Glycine max* plant between Satt168 and Satt560; and

(B) determining the probability that said plant has the quantitative trait allele for enhanced yield.

62. A method for determining a genomic polymorphism in a plant that is predictive of an enhanced yield comprising the steps:

(A) incubating a marker nucleic acid molecule, under conditions permitting nucleic acid hybridization, and a complementary nucleic acid molecule obtained from said plant, said marker nucleic acid molecule selected from the group consisting of Satt560 or its complement, Satt534 or its complement, Satt066 or its complement, U3944117 or its complement, Satt020 or its complement, Satt272 or its complement,

Sct_094 or its complement, Satt556 or its complement, Satt122 or its complement, Satt474 or its complement, Sat_083 or its complement, Satt416 or its complement, and Satt168 or its complement;

(B) permitting hybridization between said marker nucleic acid molecule and said complementary nucleic acid molecule obtained from said plant; and

(C) detecting the presence of said polymorphism.

63. A method of determining an association between a polymorphism and a plant trait comprising:

(A) hybridizing a nucleic acid molecule specific for the polymorphism to genetic material of a plant, wherein the nucleic acid molecule or complement thereof is selected from the group consisting of a nucleic acid molecule that is complementary to a nucleic acid sequence that is genetically linked to a quantitative trait locus in a region between and including a nucleic acid sequence that specifically hybridizes to a region between Satt168 and Satt560; and

(B) calculating the degree of association between the polymorphism and the plant trait.

64. A *Glycine max* plant or part thereof selected from the group consisting of C83-1, C83-2, C83-3 and C83-75 or progeny thereof.

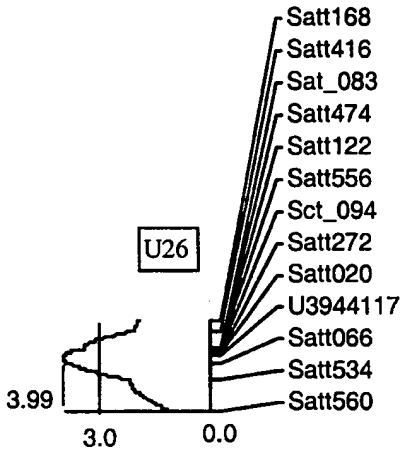
Satt168
Satt416
Sat_083
Satt474
Satt122
Satt556
Sct_094
U26
Satt272
Satt020
U3944117
Satt066
Satt534
Satt560
3.99
3.0
0.0

Figure 1.

SEQUENCE LISTING

<110> Delannay, Xavier
Concibido, Vergel C.

<120> Methods for Breeding for and Screening of Soybean Plants with Enhanced Yields, and Soybean Plants with Yields

<130> 04983.0104.PC01/38-21(15783)A PCT

<140> PCT/US99/22675
<141> 1999-09-30

<160> 40

<170> Microsoft Word 97

<210> 1
<211> 81
<212> DNA
<213> Glycine Sp.

<400> 1
gactgcgtac caattcagaa gttataagtt gtcataaata tgaatcagtt tcactctgtg 60
acaatgatgg ttccctggat t 81

<210> 2
<211> 81
<212> DNA
<213> Glycine Sp.

<400> 2
gactgcgtac caattcagaa gttataagtt gtcataaata tgcatcagtt tcactctatg 60
ataatgatgg ttccatggat t 81

<210> 3
<211> 227
<212> DNA
<213> Glycine Sp.

<400> 3
cgcttgccca aaaattaata gtatgatcat atgtggttgg agggtgggaa aaaaaagaag 60
acaataccat ataataataa taataataat aataataata ataataataa taataataat 120
aagagaaaat acatcaatac taagaagtta ttaatttaaa tgatactgaa tttaatatcc 180
ttaanttaat tctccnaaag anatataaga ttgaggttgg agaatgg 227

<210> 4
<211> 357
<212> DNA
<213> Glycine Sp.

<400> 4
tatagcccag caaaaaaaaa cagagattaa gagaagacga gagttttaag agaataagaa 60
aaattttgtg tattctttct aagaagaata aattatttat aaatacaaaa tataattgaa 120
aaaatataaa aaataaataa taataataat aataataata ataataataa taataataat 180

aataataata ataataataa taataataat aataataata ataataacaa acaatcataa 240
aagataatta aagatatgaa caatcacaca gataaattac ccataacaaa tattactaaa 300
atactaaaat tatgttatta atataaccga ttttttttgtt cattggtcgg ttttgat 357

<210> 5
<211> 252
<212> DNA
<213> Glycine Sp.

<400> 5
accattggaa tgttctacaa ttaagaataa aatctttaac attaggaaaa aaatataaaa 60
aatattaaat aaatattaaa taaaactaaa actaaaaata ttaatttaaa taaaattttt 120
tactatttta aagtgggaat atatatatat atatatatat atatatatat atatatatat 180
atatatnnnt ttgggacttc aaaatcatat tttaaaaaaa ttaagaagat gtaaactttt 240
ttataacttc aa 252

<210> 6
<211> 253
<212> DNA
<213> Glycine Sp.

<400> 6
aatttggaaa tgacatctta gaaattattc tcacaactct tgtaatatac aagtattaag 60
aatgatttta ctaacttaat aaaattcaga aaaaataata ataataataa taataataat 120
aataataata ataaaagatt gggaaggcag atagaagaat atatatgttc tcaccgcaat 180
acttggtcgt tttgtaatat ttgtagccca acatatagca gtatctcttt tcttcacatc 240
caatttctcc cgt 253

<210> 7
<211> 126
<212> DNA
<213> Glycine Sp.

<400> 7
aaccaacttg ggaatagaca ataattcaag aaatacaagt gcaagaaaga cctaataata 60
ataataataa taataatcct aaaaatggag ttaatgtctt ggtatgatta gtgaatgata 120
gagagc 126

<210> 8
<211> 156
<212> DNA
<213> Glycine Sp.

<400> 8
ataaaacccg ataaataaga tttcatgaat aataataata ataataataa taataataat 60
aataataata ctcaaagacc aaaatttcat tttcgaaaca tgatataggc ttcagagatg 120
aacgaacata aaatacataa gaaaacaagg tgcaca 156

<210> 9
<211> 114
<212> DNA
<213> Glycine Sp.

<400> 9
gggtgaagtg agagtaacac gtaagagtnc ctctctctct ctctctctct ctctctctct 60

ctctctctct ctaatactag ggggaagtta tgtctaccaa tgaagagatc cggg 114

<210> 10
<211> 221
<212> DNA
<213> Glycine Sp.

<400> 10
atgacaagga aaaatcaatc aacaatctga acctttttcc actttttttcc ttgttcaaca 60
ttataaggtt gctcacatat tatataaaga tttcatgttc tctcactcac ctcatacacc 120
attcctactc tcttaaacac acacatactt tataataata ataataataa taataataat 180
aataataata ataaaccatt ccaaacactc ttaacagcag c 221

<210> 11
<211> 115
<212> DNA
<213> Glycine Sp.

<400> 11
gagaaagaaa tgtgttagtg taataaaaaa gactaaaata ttattattat tattattatt 60
attattatta ttattattat tattatttgt gcaatcaaac aataagaagg aaaag 115

<210> 12
<211> 213
<212> DNA
<213> Glycine Sp.

<400> 12
tnnncnacgc cgcttgataa aaacacaaat ttataataat caaaaacata tttaagctta 60
ataatgaaaa tgacaccatt aaataataat aataataata ataataataa taataataat 120
aataataata ataatantaa taataataat aataataata ataatcacaa caaaaatagn 180
tcatgtaaaa tggaatgtta cagaagtgat caa 213

<210> 13
<211> 185
<212> DNA
<213> Glycine Sp.

<400> 13
ctcctcctgc gcaacaacaa tattcatgca tatacatcac gtattattat tattattatt 60
attattatta ttattattat tattattatt attattatta ttattattat tattattatt 120
attattatta taaaatgcag gtaagacatt caacaagata actaaggtca tggcctagat 180
ccccc 185

<210> 14
<211> 278
<212> DNA
<213> Glycine Sp.

<400> 14
atcgtgcaag aaaataaatt tttgaaaata attaatattt taatttttaa tatgatatta 60
attatttttt atttatattt tttatatatt actagtaaac aaaatttaaa aataaattaa 120
taataataaa atattaattt tatctatata ttattattat tattattatt attattatta 180
ttattatcat tattattatt aatagtgtgc aaaacaagtt attgtaataa gataattatt 240
tagagacgga tgaagtaatt atttgaggcg aagtccac 278

<210> 15
<211> 19
<212> DNA
<213> Glycine Sp.

<400> 15
gactgcgtac caattcaga 19

<210> 16
<211> 19
<212> DNA
<213> Glycine Sp.

<400> 16
gatgagtcct gagtaaatc 19

<210> 17
<211> 23
<212> DNA
<213> Glycine Sp.

<400> 17
cgcttgccca aaaattaata gta 23

<210> 18
<211> 25
<212> DNA
<213> Glycine Sp.

<400> 18
ccattctcca acctcaatct tatat 25

<210> 19
<211> 27
<212> DNA
<213> Glycine Sp.

<400> 19
tatagcccag caaaaaaaaa cagagat 27

<210> 20
<211> 27
<212> DNA
<213> Glycine Sp.

<400> 20
atcaaaaccg accaatgaac aaaaaaa 27

<210> 21
<211> 19
<212> DNA
<213> Glycine Sp.

<400> 21

accattggaa tgttctaca 19

<210> 22
<211> 25
<212> DNA
<213> Glycine Sp.

<400> 22
ttgaagttat aaaaaagttt acatc 25

<210> 23
<211> 27
<212> DNA
<213> Glycine Sp.

<400> 23
gcgaaatttg gaaatgacat cttagaa 27

<210> 24
<211> 27
<212> DNA
<213> Glycine Sp.

<400> 24
gcgacgggag aaattggatg tgaagaa 27

<210> 25
<211> 19
<212> DNA
<213> Glycine Sp.

<400> 25
aaccaacttg ggaatagac 19

<210> 26
<211> 22
<212> DNA
<213> Glycine Sp.

<400> 26
gctctctatc attcactaat ca 22

<210> 27
<211> 21
<212> DNA
<213> Glycine Sp.

<400> 27
gcgataaaac ccgataaata a 21

<210> 28
<211> 21
<212> DNA
<213> Glycine Sp.

<400> 28
gcgttgtgca ccttgttttc t 21

<210> 29
<211> 19
<212> DNA
<213> Glycine Sp.

<400> 29
gggtgaagtg agagtaaca 19

<210> 30
<211> 16
<212> DNA
<213> Glycine Sp.

<400> 30
cccggatctc ttcatt 16

<210> 31
<211> 23
<212> DNA
<213> Glycine Sp.

<400> 31
atgacaagga aaaatcaatc aac 23

<210> 32
<211> 20
<212> DNA
<213> Glycine Sp.

<400> 32
gctgctgtta agagtgtttg 20

<210> 33
<211> 20
<212> DNA
<213> Glycine Sp.

<400> 33
tttgaaggaa gggtggtgag 20

<210> 34
<211> 22
<212> DNA
<213> Glycine Sp.

<400> 34
gatccaaatc ctcagtatca ta 22

<210> 35
<211> 26
<212> DNA
<213> Glycine Sp.

<400> 35
gggaagctta ataatgaaaa tgacac 26

<210> 36
<211> 21
<212> DNA
<213> Glycine Sp.

<400> 36
ttgatcactt ctgtaacatt c 21

<210> 37
<211> 22
<212> DNA
<213> Glycine Sp.

<400> 37
ctcctcctgc gcaacaacaa ta 22

<210> 38
<211> 19
<212> DNA
<213> Glycine Sp.

<400> 38
gggggatcta ggccatgac 19

<210> 39
<211> 19
<212> DNA
<213> Glycine Sp.

<400> 39
gcgatcgtgc aagaaaata 19

<210> 40
<211> 24
<212> DNA
<213> Glycine Sp.

<400> 40
gcggtggact tcgcctcaaa taat 24

INTERNATIONAL SEARCH REPORT

Int. ional Application No
PCT/US 99/22675

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68 C12N15/82 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 746 023 A (HANAFEY MICHAEL K ET AL) 5 May 1998 (1998-05-05) the whole document ---	
A	KEIM P ET AL: "RFLP MAPPING IN SOYBEAN: ASSOCIATION BETWEEN MARKER LOCI AND VARIATION IN QUANTITATIVE TRAITS" GENETICS,US,HILL, NC, vol. 126, no. 3, 1 November 1990 (1990-11-01), pages 735-742, XP000604563 the whole document ---	

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 February 2000	15/02/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Knehr, M

3

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int. ional Application No
PCT/US 99/22675

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication,where appropriate. of the relevant passages	Relevant to claim No.
A	MANSUR L M ET AL.: "Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (Glycine max L.)" THEORETICAL AND APPLIED GENETICS, vol. 86, 1993, pages 907-913, XP000874256 the whole document	
A	ANKSLEY S D AND NELSON J C: "Avanced backcross QTL analysis: a method for the simultanous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines" THEORETICAL AND APPLIED GENETICS, vol. 92, 1996, pages 191-203, XP000874433 cited in the application the whole document	

3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 99/ 22675

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-13,25-37,39-64
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 1-13, 25-37, 39-64

Present claims 1-13, 25-37, 39-64 refer essentially to glycine plants, seeds, nucleic acid molecules, and methods of production/screening, basically relying on quantitative trait loci of linkage group U26 (and markers localized within that linkage group) associated with enhanced yield. However, the most relevant features within these claims are represented by linkage group U26, as well as glycine strains PI407305, C83-1, C83-2, C83-3, and C83-75, remain completely undefined and are not supported by any technical feature within the description, so therefore, a meaningful search is not possible.
Thus, related to a significant part of the entire concept of the application as disclosed and claimed, it is impossible to determine the matter for which protection is sought. Accordingly, these parts of the present application fail to comply with the requirement of article 6 PCT, first sentence 'see also rule 6.1(a) PCT! and fail to comply with the requirements for clarity and conciseness of article 6 PCT, second sentence.
Consequently, the search has been limited to and carried out for those parts of the application which do appear to be clear and concise, namely claims 14-24, and 38.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int tional Application No
PCT/US 99/22675

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5746023 A	05-05-1998	US 5437697 A	01-08-1995

CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12Q 1/68, C12N 15/82, A01H 5/00

A1

(11) International Publication Number: WO 00/18963

(43) International Publication Date: 6 April 2000 (06.04.00)

(21) International Application Number: PCT/US99/22675

(22) International Filing Date: 30 September 1999 (30.09.99)

(30) Priority Data:

60/102,662	1 October 1998 (01.10.98)	US
60/127,627	1 April 1999 (01.04.99)	US
60/135,608	24 May 1999 (24.05.99)	US

(71) Applicant: MONSANTO COMPANY [US/US]; 800 North Lindbergh Boulevard, St. Louis, MO 63167 (US).

(72) Inventors: DELANNAY, Xavier; 13143 Four Poster Court, St. Louis, MO 63146 (US). CONCIBIDO, Vergel, C.; 379 Birchwood Crossing Lane, Maryland Heights, MO 63043 (US).

(74) Agent: MARSH, David, R.; Howrey & Simon, Box 34, 1299 Pennsylvania Avenue, N.W., Washington, DC 20004-2402 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS FOR BREEDING FOR AND SCREENING OF SOYBEAN PLANTS WITH ENHANCED YIELDS, AND SOYBEAN PLANTS WITH ENHANCED YIELDS

(57) Abstract

The present invention is in the field of plant breeding and genetics, particularly as it pertains to *Glycine max* (soybean). More specifically, the invention relates to quantitative trait loci that are associated with enhanced yield in *Glycine max*, *Glycine max* having such loci and methods for breeding for and screening of *Glycine max* with such loci. The invention further relates to the use of exotic germplasm in a breeding program.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		